

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKYAAFAIVVSGSALA.....VTRVTHMAHANATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 779 | 100.0 | 151 | 3 AAB36346 | Aab36346 Agfa::PT3 |
| 2 | 696 | 89.3 | 151 | 3 AAB36347 | Aab36347 Agfa::PT3 |
| 3 | 695 | 89.2 | 151 | 2 AAR74625 | Aar74625 Agfa sequ |
| 4 | 695 | 89.2 | 151 | 3 AAB36341 | Aab36341 Salmonell |
| 5 | 690 | 88.6 | 151 | 2 AAW23570 | Aaw23570 Salmonell |
| 6 | 655 | 84.1 | 151 | 3 AAB36352 | Aab36352 Agfa::PT3 |
| 7 | 614 | 78.8 | 151 | 3 AAB36353 | Aab36353 Agfa::PT3 |
| 8 | 613 | 78.7 | 151 | 3 AAB36349 | Aab36349 Agfa::PT3 |
| 9 | 611 | 78.4 | 151 | 3 AAB36350 | Aab36350 Agfa::PT3 |
| 10 | 605 | 77.7 | 151 | 3 AAB36354 | Aab36354 Agfa::PT3 |
| 11 | 604 | 77.5 | 151 | 3 AAB36351 | Aab36351 Agfa::PT3 |
| 12 | 603 | 77.4 | 151 | 3 AAB36355 | Aab36355 Agfa::PT3 |
| 13 | 580 | 74.5 | 151 | 3 AAB36348 | Aab36348 Agfa::PT3 |
| 14 | 560 | 71.9 | 120 | 2 AAR62761 | Aar62761 Agfa sequ |
| 15 | 560 | 71.9 | 120 | 2 AAW23569 | Aaw23569 Salmonell |
| 16 | 520 | 66.8 | 151 | 3 AAB36343 | Aab36343 Escherich |
| 17 | 515 | 66.1 | 151 | 7 ABR82651 | AbR82651 E. coli C |
| 18 | 485 | 62.3 | 142 | 2 AAR52664 | Aar52664 Fibronect |
| 19 | 413 | 53.0 | 122 | 2 AAR52663 | Aar52663 FNB curli |
| 20 | 237 | 30.4 | 45 | 3 AAB36316 | Aab36316 Salmonell |
| 21 | 132 | 16.9 | 22 | 3 AAB36318 | Aab36318 Salmonell |
| 22 | 123 | 15.8 | 23 | 3 AAB36321 | Aab36321 Salmonell |
| 23 | 123 | 15.8 | 23 | 3 AAB36326 | Aab36326 Salmonell |
| 24 | 123 | 15.8 | 23 | 3 AAB36338 | Aab36338 Salmonell |
| 25 | 115 | 14.8 | 22 | 3 AAB36325 | Aab36325 Salmonell |

| | | | | | |
|----|-------|------|------|------------|--------------------|
| 26 | 115 | 14.8 | 22 | 3 AAB36339 | Aab36339 Salmonell |
| 27 | 115 | 14.8 | 22 | 3 AAB36320 | Aab36320 Salmonell |
| 28 | 113 | 14.5 | 24 | 7 ABR82644 | AbR82644 E. coli C |
| 29 | 109 | 14.0 | 23 | 3 AAB36340 | Aab36340 Salmonell |
| 30 | 109 | 14.0 | 23 | 3 AAB36324 | Aab36324 Salmonell |
| 31 | 109 | 14.0 | 23 | 3 AAB36319 | Aab36319 Salmonell |
| 32 | 100.5 | 12.9 | 151 | 3 AAB36342 | Aab36342 Salmonell |
| 33 | 98.5 | 12.6 | 850 | 4 ABB65764 | Abb65764 Drosophil |
| 34 | 98.5 | 12.6 | 1028 | 4 ABB62708 | Abb62708 Drosophil |
| 35 | 98 | 12.6 | 26 | 7 ABR82645 | AbR82645 E. coli C |
| 36 | 95 | 12.2 | 24 | 7 ABR82647 | AbR82647 E. coli C |
| 37 | 94.5 | 12.1 | 678 | 6 ABU36649 | AbU36649 Protein e |
| 38 | 94.5 | 12.1 | 688 | 5 ABR74039 | AbP74039 Candida a |
| 39 | 93 | 11.9 | 764 | 6 AAB36890 | Aae36890 Plectreir |
| 40 | 92 | 11.8 | 23 | 3 AAB36331 | Aae36331 Escherich |
| 41 | 92 | 11.8 | 673 | 3 AAY44403 | Aay44403 Human tru |
| 42 | 92 | 11.8 | 673 | 5 AAU79538 | Aau79538 Truncated |
| 43 | 92 | 11.8 | 949 | 3 AAY44404 | Aay44404 Human tru |
| 44 | 92 | 11.8 | 949 | 5 AAU79539 | Aau79539 Truncated |
| 45 | 92 | 11.8 | 1327 | 3 AAY44402 | Aay44402 Human tan |

ALIGNMENTS

RESULT 1

AAB36346
ID AAB36346 standard; protein; 151 AA.
XX
AC AAB36346;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64622.
XX
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEPI7/TA) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA, and Agfa-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 779; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2e-66; Mismatches 0; Gaps 0;
 Matches 151; Conservative 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNTANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNTANQY 151

RESULT 2

AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN W0200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX

SQ Sequence 151 AA;

Query Match 89.3%; Score 696; DB 3; Length 151;
 Best Local Similarity 87.6%; Pred. No. 1.7e-58;
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 117
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNTANQY 151
 DB 118 -----YDQLVTRVVTHEMAHANNTANQY 151

RESULT 3

AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 DE Salmonella; Agfa; vaccine.
 KW Salmonella.
 OS Salmonella.
 XX
 PN W09425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 PR 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 151 AA;
 Query Match 89.2%; Score 695; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGDPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNDQLVTRVVTTHMAHANNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 DR WPI; 2000-672631/55.
 DR N-PSDB; AAC64617.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CSga and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Query Match 89.2%; Score 695; DB 3; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGDPSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGDPSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNDQLVTRVVTTHMAHANNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;

```

XX WPI: 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfa encoded by the full agfa gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteriaceae. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridize to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
SQ
Query Match      88.6%; Score 690; DB 2; Length 151;
Best Local Similarity 89.4%; Pred. No. 6.5e-58;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 NNPALVNYQTASDSSVMVRQVGFNNATANQY 151
RESULT 6
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
XX 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64628.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
Query Match      84.1%; Score 655; DB 3; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.4e-54;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151
DB 110 -----NYDQLVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
RESULT 7
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
XX 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
XX
XX

```


AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 PI
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 137; 139pp; English.
 XX
 XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC useful for the expression of a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 78.4%; Score 611; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.3e-50;
 Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQGSANAALALQ 60
 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGQCADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKYDQLVTRVVTTHMAHAGQAGDNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNGNATANQY 151
 RESULT 10
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX
 AC AAB36354;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 XX AgfA::PT3#9 amino acid sequence SEQ ID NO:28.
 DE
 XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 PI
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64630.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 138; 139pp; English.
 XX
 XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC useful for the expression of a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 78.4%; Score 611; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.3e-50;
 Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQGSANAALALQ 60
 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQGSANAALALQ 60

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.7%; Score 605; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 8.5e-50;
 Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGYGG 120
 QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151

RESULT 11
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 XX
 AC AAB36351;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 FA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX

The present invention describes a recombinant agfa gene (I) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fimbriae (SPF17/TAP) nucleation depended
 assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 77.5%; Score 604; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.1e-49;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDIIVGYGG 120
 QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151

RESULT 12
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 XX
 AC AAB36355;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 FA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64631.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 139; 139pp; English.
 XX

The present invention describes a recombinant agfa gene (I) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fimbriae (SPF17/TAP) nucleation depended
 assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 77.4%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 1.3e-49;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX AAB36348;
 AC AAB36348;
 XX 26-FEB-2001 (first entry)
 DT DT
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 PN 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 PR (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 XX 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 74.5%; Score 580; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 2.1e-47;
 Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 14
 AAR62761
 ID AAR62761 standard; protein; 120 AA.
 XX AAR62761;
 AC AAR62761;
 XX 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX AgfA sequence.
 DE Salmonella; AgfA; vaccine.
 XX Salmonella enteritidis.
 OS WO9425598-A2.
 PN 10-NOV-1994.

```

XX 26-APR-1994; 94WO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX
XX WPI; 1994-358275/44.
XX N-PSDB; AAQ73066.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
XX strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The sequence represents the Salmonella enteritis 27655-3b TnpHoA mutant
XX strain Agfa protein. The encoding DNA and isolated Agfa protein are used
XX in genetic immunization and vaccine compositions, respectively, to elicit
XX an immune response to Salmonella in animals (e.g. food producing animals)
XX and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
XX AUG-2003 to correct OS field.)
XX
XX Sequence 120 AA;
SQ
Query Match 71.9%; Score 560; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 VVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
DB 1 VVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 127
DB 61 GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 106
RESULT 15
AAW23569
ID AAW23569 standard; protein; 120 AA.
XX
XX AAW23569;
XX
XX 25-MAR-2003 (revised)
XX 29-SEP-1997 (first entry)
XX
XX Salmonella enteritis 27655-3b TnpHoA mutant agfa fragment.
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
XX Salmonella enteritidis.
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX Collinson SK, Kay WW, Doran JL;
XX
XX WPI; 1997-309886/28.
XX N-PSDB; AAR74141.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX

```

```

PS Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents an agfa fragment encoded by an agfa gene
XX fragment derived from Salmonella enteritidis 27655-3b TnpHoA mutant
XX strain. The nucleic acid can be used to provide diagnostic assays for
XX Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.
XX It can also be used to provide proteins and antibodies which can be used
XX for assays. The nucleic acid sequence can be used to provide probes or
XX primers which can specifically hybridise to nucleic acid molecules from
XX greater than 99% of Salmonella strains that are pathogenic to warm-
XX blooded animals relative to nucleic acid molecules from virtually all
XX other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 120 AA;
SQ
Query Match 71.9%; Score 560; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 22 VVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
DB 1 VVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 127
DB 61 GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYGGNNAALVN 106
Search completed: August 2, 2004, 14:48:24
Job time : 45.9 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVTHMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 690 | 88.6 | 151 | 1 | US-08-233-788A-59 |
| 2 | 560 | 71.9 | 120 | 1 | US-08-233-788A-57 |
| 3 | 92 | 11.8 | 673 | 3 | US-09-196-387-8 |
| 4 | 92 | 11.8 | 673 | 4 | US-09-841-835-8 |
| 5 | 92 | 11.8 | 949 | 3 | US-09-196-387-10 |
| 6 | 92 | 11.8 | 949 | 4 | US-09-841-835-10 |
| 7 | 92 | 11.8 | 1327 | 3 | US-09-196-387-2 |
| 8 | 92 | 11.8 | 1327 | 4 | US-09-841-835-2 |
| 9 | 92 | 11.8 | 1327 | 4 | US-09-972-115A-8 |
| 10 | 90.5 | 11.6 | 745 | 4 | US-09-336-115C-6 |
| 11 | 89.5 | 11.5 | 738 | 3 | US-08-864-038A-3 |
| 12 | 89 | 11.4 | 943 | 4 | US-09-056-556-204 |
| 13 | 89 | 11.4 | 943 | 4 | US-09-072-596-199 |
| 14 | 89 | 11.4 | 943 | 4 | US-09-477-135A-131 |
| 15 | 89 | 11.4 | 943 | 4 | US-09-072-967-204 |
| 16 | 87.5 | 11.2 | 892 | 4 | US-09-336-447A-5 |
| 17 | 85 | 10.9 | 956 | 4 | US-09-134-078-63 |
| 18 | 82.5 | 10.6 | 186 | 4 | US-09-382-276-3 |
| 19 | 82.5 | 10.6 | 873 | 4 | US-09-336-447A-13 |
| 20 | 81.5 | 10.5 | 339 | 4 | US-09-252-991A-32096 |
| 21 | 81 | 10.4 | 568 | 4 | US-09-543-681A-6965 |
| 22 | 81 | 10.4 | 1739 | 4 | US-09-540-236-3739 |
| 23 | 81 | 10.4 | 1864 | 2 | US-08-804-227C-3 |
| 24 | 80 | 10.3 | 232 | 4 | US-09-252-921A-30263 |
| 25 | 80 | 10.3 | 975 | 4 | US-09-328-352-4764 |
| 26 | 80 | 10.3 | 3241 | 4 | US-09-841-786-1 |
| 27 | 79.5 | 10.2 | 361 | 4 | US-09-540-236-2164 |

28 79 10.1 941 4 US-09-336-447A-9 Sequence 9, Appli
29 78.5 10.1 2123 3 US-08-968-685A-10 Sequence 10, Appli
30 78 10.0 306 1 US-08-254-573-2 Sequence 2, Appli
31 78 10.0 306 1 US-08-687-379-4 Sequence 2, Appli
32 78 10.0 306 1 US-08-687-379-4 Sequence 4, Appli
33 78 10.0 306 4 US-08-172-332-1 Sequence 1, Appli
34 78 10.0 306 4 US-08-216-326-2 Sequence 2, Appli
35 77.5 9.9 702 4 US-09-252-991A-22119 Sequence 22119, A
36 77.5 9.9 714 4 US-09-841-786-4 Sequence 4, Appli
37 77.5 9.9 878 4 US-09-540-236-3401 Sequence 3401, Ap
38 77 9.9 415 4 US-09-025-769B-280 Sequence 280, App
39 77 9.9 528 4 US-09-490-291-8 Sequence 8, Appli
40 77 9.9 1690 4 US-09-595-684B-39 Sequence 39, Appli
41 77 9.9 2315 4 US-09-543-681A-5434 Sequence 5434, Ap
42 76.5 9.8 159 3 US-08-856-253-2 Sequence 2, Appli
43 76.5 9.8 186 4 US-09-382-276-2 Sequence 2, Appli
44 76.5 9.8 605 4 US-09-489-039A-13002 Sequence 13002, A
45 75.5 9.7 624 4 US-09-336-447A-7 Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617

; GENERAL INFORMATION:

; APPLICANT: Doran, James L.

; APPLICANT: Key, William W.

; APPLICANT: Collinson, Karen S.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

; TITLE OF INVENTION: OF SALMONELLA

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,788A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 59:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 151 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: Protein

; US-08-233-788A-59

Query Match 88.6%; Score 690; DB 1; Length 151;

Best Local Similarity 89.4%; Pred. No. 7, 1e-62;

Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALGVVPPQWGGNGNGSGPDSITLIYQVGSANALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SPARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151
Db 121 NNPALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Key, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 71.9%; Score 560; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 6e-49;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81
Db 1 VVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVN 127
Db 61 GQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVN 106

RESULT 3
US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-196-387-8

Query Match 11.8%; Score 92; DB 3; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.6;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSPSSSSPTSS-SSSSPSPGSSLAESPAA 157
QY 65 KSETTIT---QSGYNGADVQGDNSTIETQNG--FRNNATIDQWNAKNSDI 113
Db 158 GVSSTAPLGPAGPGTGVPAVSGALRELLEACRNGDVSRLVDAAVNAKDM 212

RESULT 4
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible


```
; SOFTWARE: PatentIn version 3.11
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match      11.8%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.5;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VYGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALQSDAR 64
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNGNPGSSSSPTSS--SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYGNAGDVGGADNSTTIELTONG--FRNNATIDOWNAKNGDI 113
DB 158 GVSSTAPLPGCAAGPGTGVPVAVSGALRELLAEACRGDVSVRKVLVDAAVNAKDM 212

RESULT 10
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336.115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match      11.6%; Score 90.5; DB 4; Length 745;
Best Local Similarity 20.8%; Pred. No. 0.97;
Matches 40; Conservative 28; Mismatches 63; Indels 61; Gaps 8;

QY 8 APAIIVYGSALAGVVPW-----GGGNGHNGGN----- 37
DB 102 AYQAVFLAINAAVGL---WNTIGYAVMGCGNGTSGPGSVIFNDPQGDSTQITCNRPE 158
QY 38 SSGPSTLSIYQGSANAALQSDARKSETTITQSGYNG-----ADVQ 83
DB 159 STPGKSNISDEFKKLNAYQLIQALKNQSGFFELG--GNTKVSVYNYECROTADNG 217
QY 84 G-----ADNSTIELTONGFRNNATIDOWNAKNSDITVQYGGNNAALVNYDQLVTRVVT 137
DB 218 GYQPCAKNGS--SSSSNGGSSSTQTATTATTDGVTITTYNNKATVKFD-----IT 270
QY 138 HEMAHANNATAN 149
DB 271 NNAEQLNQAN 282

RESULT 11
US-08-864-038A-3
; Sequence 3, Application US/08864038A
```

```
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match      11.5%; Score 89.5; DB 3; Length 738;
Best Local Similarity 35.4%; Pred. No. 1.2;
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;

QY 3 LKVAFAAIIVYGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALQSD 62
DB 419 LKSSASASASASASAG-----GGGGGNGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYGNAGDVGG 84
DB 461 -----AALAAAGAGGGLGGGG 477

RESULT 12
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-056-556-204

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQYGSANAALALQSDAR---KSET 68
Db 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFNSGTGNGVIGNSGTGNGWIGNSG 205

QY 69 TITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGQY----- 118
Db 206 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNYTGSNPGNSNTGPFNMGOYNTGYLNS 265

QY 119 GGNNAALVN 127
Db 266 GNYNTGLAN 274

RESULT 13
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dallon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 350
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-199

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQYGSANAALALQSDAR---KSET 68
Db 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFNSGTGNGVIGNSGTGNGWIGNSG 205

QY 69 TITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGQY----- 118
Db 206 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNYTGSNPGNSNTGPFNMGOYNTGYLNS 265

QY 119 GGNNAALVN 127
Db 266 GNYNTGLAN 274

RESULT 14
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Naro, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-477-135A-131

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;

QY 12 IVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQYGSANAALALQSDAR---KSET 68
Db 266 IGLTSGLLGF-----GGLNSGTGN-----IGLFNSGTGNGVIGNSGTGNGWIGNSG 312

QY 69 TITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD---ITVGQY----- 118

```

Job time : 13 secs

Db 313 NSYNTGFGNSGDANTGFFNSGIAN TGVCNAGNYNTGTSYNPGNSNTGFGNMGQYNTGYLNS 372

QY 119 GGNNAALVN 127
Db 373 GNYNTGLAN 381

RESULT 15

US-09-072-967-204
; Sequence 204, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-967-204

Query Match 11.4%; Score 89; DB 4; Length 943;
Best Local Similarity 25.6%; Pred. No. 1.9;
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;
QY 12 IVVSGSALAGVVPQGGNHNHGGNCGSSGPDSTLSIYQYGSANAALQSDAR---KSET 68
Db 159 IGLTSGGLGF-----GGLNSGTGN-----IGLNSGTGNVIGNSGTGNWIGNSG 205
QY 69 TTQSGYGNAGDVGGADNSTIELTQNGFRNNAIDOWNAKNSD---ITVQY----- 118
Db 206 NSYNTGFGNSGDANTGFFNSGIAN TGVCNAGNYNTGTSYNPGNSNTGFGNMGQYNTGYLNS 265
QY 119 GGNNAALVN 127
Db 266 GNYNTGLAN 274

Search completed: August 2, 2004, 14:58:31

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLKVAFAAIVVSGSALA.....VTRVTHMAHANNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 517 | 66.4 | 151 | 12 | US-09-741-873B-4 |
| 2 | 517 | 66.4 | 151 | 12 | US-09-741-873B-4 |
| 3 | 439 | 56.4 | 131 | 12 | US-09-741-873B-2 |
| 4 | 439 | 56.4 | 131 | 12 | US-09-741-873B-2 |
| 5 | 100 | 12.8 | 445 | 15 | US-10-369-493-20638 |
| 6 | 97.5 | 12.5 | 263 | 12 | US-10-425-114-49960 |
| 7 | 94.5 | 12.1 | 678 | 12 | US-10-282-122A-64573 |
| 8 | 94.5 | 12.1 | 688 | 14 | US-10-032-585-7876 |
| 9 | 93 | 11.9 | 271 | 16 | US-10-437-963-147343 |
| 10 | 92 | 11.8 | 673 | 9 | US-09-841-835-8 |
| 11 | 92 | 11.8 | 949 | 9 | US-09-841-835-10 |
| 12 | 92 | 11.8 | 1327 | 9 | US-09-841-835-2 |
| 13 | 92 | 11.8 | 1327 | 10 | US-09-972-115A-8 |
| 14 | 92 | 11.8 | 1327 | 14 | US-10-199-937-4 |
| 15 | 91.5 | 11.7 | 705 | 16 | US-10-437-963-203823 |

| | | | | | |
|----|------|------|------|----|----------------------|
| 16 | 91 | 11.7 | 154 | 16 | US-10-437-963-162284 |
| 17 | 90.5 | 11.6 | 745 | 8 | US-08-834-666A-6 |
| 18 | 90.5 | 11.6 | 745 | 12 | US-10-282-122A-58683 |
| 19 | 90 | 11.6 | 1778 | 14 | US-10-238-075-749 |
| 20 | 89.5 | 11.5 | 270 | 16 | US-10-437-963-122263 |
| 21 | 89.5 | 11.5 | 486 | 15 | US-10-369-493-20619 |
| 22 | 89 | 11.4 | 943 | 9 | US-09-996-634-131 |
| 23 | 89 | 11.4 | 943 | 10 | US-09-997-182-131 |
| 24 | 89 | 11.4 | 943 | 10 | US-09-997-181-131 |
| 25 | 89 | 11.4 | 943 | 14 | US-10-193-002-199 |
| 26 | 89 | 11.4 | 943 | 14 | US-10-084-843-204 |
| 27 | 89 | 11.4 | 3300 | 12 | US-10-282-122A-64369 |
| 28 | 88.5 | 11.4 | 145 | 16 | US-10-437-963-147748 |
| 29 | 88.5 | 11.4 | 191 | 16 | US-10-437-963-105413 |
| 30 | 88 | 11.3 | 354 | 10 | US-09-820-843A-21 |
| 31 | 88 | 11.3 | 616 | 16 | US-10-437-963-193067 |
| 32 | 87.5 | 11.2 | 253 | 16 | US-10-437-963-114193 |
| 33 | 87.5 | 11.2 | 628 | 12 | US-10-282-122A-53269 |
| 34 | 87.5 | 11.2 | 892 | 10 | US-09-952-287-5 |
| 35 | 87 | 11.2 | 482 | 14 | US-10-156-761-8763 |
| 36 | 86 | 11.0 | 448 | 16 | US-10-437-963-140685 |
| 37 | 86 | 11.0 | 2457 | 12 | US-10-282-122A-49854 |
| 38 | 85.5 | 11.0 | 191 | 16 | US-10-437-963-152921 |
| 39 | 85.5 | 11.0 | 942 | 12 | US-10-282-122A-73345 |
| 40 | 85.5 | 11.0 | 1862 | 12 | US-10-282-122A-49757 |
| 41 | 85 | 10.9 | 257 | 10 | US-09-880-748-1494 |
| 42 | 85 | 10.9 | 257 | 12 | US-10-293-418-1494 |
| 43 | 85 | 10.9 | 956 | 13 | US-10-121-032-63 |
| 44 | 85 | 10.9 | 956 | 14 | US-10-093-037-63 |
| 45 | 84.5 | 10.8 | 271 | 14 | US-10-156-761-11721 |

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli

Query Match 66.4%; Score 517; DB 12; Length 151;
Best Local Similarity 66.3%; Pred. No. 1.7e-43;
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLKVAFAAIVVSGSALAGVTFQGGGNGHNGGNSGPDSTLSIYQVGSANALALQ 60
US-09-741-873B-4

```
Db 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
Db 61 TDARNSDLTITQGGNGADVGCGSDSSIDLQTFQFGNSATLDQNGKNSMTVKQFGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 2

```
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4
```

```
Query Match 66.4%; Score 517; DB 12; Length 151;
Best Local Similarity 66.2%; Pred. No. 1.7e-43;
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGSANAALALQ 60
Db 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
Db 61 TDARNSDLTITQGGNGADVGCGSDSSIDLQTFQFGNSATLDQNGKNSMTVKQFGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 3

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 56.4%; Score 439; DB 12; Length 131;
Best Local Similarity 62.6%; Pred. No. 8e-36;
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHGGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQYGGGNGHGGGNSGPNSELNTYQYGGNSALALQTDARNSDLTITQHGNGGAD 60
QY 81 VGOAGDNSTIELTQNGFRNATIDOWNAKNSDITVGOYGNNAALVNYDQVTRVVTHEM 140
Db 61 VQGSDDSSIDLQTFQFGNSATLDQNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQ 120
QY 141 AHANNATANY 151
Db 121 GFGNNATAHQY 131
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 56.4%; Score 439; DB 12; Length 131;
```



```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64573
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64573

Query Match      12.1%; Score 94.5; DB 12; Length 678;
Best Local Similarity 25.7%; Pred. No. 1.1;
Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;

QY 15 SCSALAGVVPQWGGGHNHGGNSGPDSTLSIYQYGSANAALALQ--SDARKSETTITQ 72
Db 384 SSGNGLG----FGSGNGNIGFFNSG--NNHGNHGGNSGNGVSGALSVFSGSAERS----- 432
QY 73 SGYNGADYVQGGADNS-----TIELTQNGFRNNATIDQ--WNAKSDITVGVGGNN 122
Db 433 SGFNSGSLSTGLGNSGQLSTGWFNSAT--STGWFNSGTTNTGWFNSGTTNTGIGNSGGN- 491
QY 123 AALVNYDQLVTRVVTHEM-----AHANNATAN 149
Db 492 -----LVTGSMGLFNSGHTNTGSPN 511

RESULT 8
US-10-032-585-7876
; Sequence 7876, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7876
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7876

Query Match      12.1%; Score 94.5; DB 14; Length 688;
Best Local Similarity 25.8%; Pred. No. 1.1;
Matches 31; Conservative 24; Mismatches 42; Indels 23; Gaps 4;

QY 30 GHHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYVQGGADNST 89
Db 488 GNNNGSGNSGTTNNNNYNN-----KSISKNE-----IDGDDLNTPTSIINN 531
QY 90 IELTQNGFRNNATIDQWNAKSDITVGVGGNNAALVNYDQLVTRVVTHEMAHANNATAN 149
Db 532 TGLTNN---NNSKSPAKSKKSNFD-----NNSNSALNLDKSKLKINTNEITN-SETTSN 584

RESULT 9
US-10-437-963-147343
; Sequence 147343, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147343
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47881C.1.pep
US-10-437-963-147343

Query Match      11.9%; Score 93; DB 16; Length 271;
Best Local Similarity 22.2%; Pred. No. 0.49;
Matches 34; Conservative 18; Mismatches 59; Indels 42; Gaps 5;

QY 14 VSGSALAGVVPQWGGG--GNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQ 72
Db 7 VVGRAAAAAYPEVSGSDGGSGGGGEGGE-----GGDGSVAAVNPEAGSGGDRSS 58
QY 73 SGYNGADYVQGGADNSTIELTQNGFRNNATIDQNA-----KNSDITVGV 118
Db 59 GEGGSGESGGGLGRQR-----YWNRLSTERQRLVDHVFKNSDVVCDFV 103

QY 119 GGNNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 104 SGVGPFAIS----AARKVYVYANDLNPTAVEY 132

RESULT 10
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-8
```

Query Match 11.8%; Score 92; DB 9; Length 673;
Best Local Similarity 30.4%; Pred. No. 1.9;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGNGHNGGNSGDPSTLSIYQVGSANAALALQSDAR 64
D5 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDADNSTIELTQNG--FRNNATIDQWNAKNSDI 113
D5 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 11

US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIPI
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.8%; Score 92; DB 9; Length 949;
Best Local Similarity 30.4%; Pred. No. 2.9;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGNGHNGGNSGDPSTLSIYQVGSANAALALQSDAR 64
D5 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDADNSTIELTQNG--FRNNATIDQWNAKNSDI 113
D5 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 12

US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIPI
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match 11.8%; Score 92; DB 9; Length 1327;
Best Local Similarity 30.4%; Pred. No. 4.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGNGHNGGNSGDPSTLSIYQVGSANAALALQSDAR 64
D5 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDADNSTIELTQNG--FRNNATIDQWNAKNSDI 113
D5 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 13

US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Publication No. US20030032769A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577

; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match 11.8%; Score 92; DB 10; Length 1327;
Best Local Similarity 30.4%; Pred. No. 4.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNSPSSSPTSS-SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVCGGADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKOM 212

RESULT 14
US-10-199-937-4
; Sequence 4, Application US/10199937
; Publication No. US20030190739A1
; GENERAL INFORMATION:
; APPLICANT: Christenson, Erik
; APPLICANT: Demaggio, Anthony J.
; APPLICANT: Goldman, Phyllis S.
; APPLICANT: McElligott, David L.
; TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS

; FILE REFERENCE: 27866/36559
; CURRENT APPLICATION NUMBER: US/10/199,937
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US/09/606,035
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/141,582
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 178
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-937-4

Query Match 11.8%; Score 92; DB 14; Length 1327;
Best Local Similarity 30.4%; Pred. No. 4.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNSPSSSPTSS-SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVCGGADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKOM 212

RESULT 15
US-10-437-963-203823
; Sequence 203823, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 203823
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_98970C.1.pap
US-10-437-963-203823

Query Match 11.7%; Score 91.5; DB 16; Length 705;
Best Local Similarity 24.7%; Pred. No. 2.3;
Matches 42; Conservative 19; Mismatches 56; Indels 53; Gaps 7;
QY 12 IVVSGSALAGVVPQWGGGNNHNGGNSGDPSTLSIYQYGSANAALALQSDARKSET 68
Db 112 IFSGGSSSAGS---GGGGGGGGSTASSPLTNALPTGNICPSARVASAAPPRRARP 167
QY 69 TITQSG---YNGG-----ADVGGADNSTIETLQNGER 98
Db 168 DVLGSGTGHYGHGSIIMRGGGMTPPRSSIDASPHYHSYRSRPAQSGSGGLQEVTRAG-- 225
QY 99 NNATIDOWNAKNSDIIVGOYGGNNAALVNYDOLVTRVVTHEMAHANNATA 148
Db 226 -----NEWYKK-----GHYG---EALRYDQVALCPDPSAACSNNRAAA 261

Search completed: August 2, 2004, 15:36:09
Job time : 37.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-12
Perfect score: 779
Sequence: 1 MLLKVAFAIAIVVGSALA.....VTRVTHMAHANNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main.*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|----|-------------|
|------------|-------|--------------------|----|-------------|

| | | | | | | |
|----|-------|-------|------|----|----------------------|-------------------|
| 1 | 779 | 100.0 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 2 | 696 | 89.3 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 3 | 690 | 89.2 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 4 | 690 | 88.6 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 5 | 655 | 84.1 | 151 | 19 | US-09-543-407-26 | Sequence 24, Appl |
| 6 | 614 | 78.8 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 7 | 613 | 78.7 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 8 | 611 | 78.4 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 9 | 608 | 78.0 | 131 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 10 | 605 | 77.7 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 11 | 604 | 77.5 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 12 | 603 | 77.4 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 13 | 580 | 74.5 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 560 | 71.9 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 15 | 520 | 66.8 | 151 | 19 | US-08-543-407-7 | Sequence 7, Appl |
| 16 | 517 | 66.4 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 17 | 517 | 66.4 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 18 | 515 | 66.1 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 19 | 515 | 66.1 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 20 | 481 | 61.7 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 21 | 481 | 61.7 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 481 | 61.7 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 23 | 476 | 61.1 | 109 | 19 | US-09-543-407-34 | Sequence 2, Appl |
| 24 | 439 | 56.4 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 439 | 56.4 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 347 | 44.5 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 27 | 335 | 43.0 | 109 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 28 | 247.5 | 31.8 | 70 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 29 | 237 | 30.4 | 48 | 19 | US-08-543-407-39 | Sequence 39, Appl |
| 30 | 100.5 | 12.9 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 31 | 100.5 | 12.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 100.5 | 12.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 33 | 100.5 | 12.9 | 186 | 30 | US-10-417-886-5833 | Sequence 5833, Ap |
| 34 | 100 | 12.8 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 35 | 100 | 12.8 | 445 | 33 | US-60-360-039-20638 | Sequence 20638, A |
| 36 | 100 | 12.8 | 1249 | 30 | US-10-455-719-358 | Sequence 358, App |
| 37 | 100 | 12.8 | 1249 | 33 | US-60-446-775-358 | Sequence 358, App |
| 38 | 100 | 12.8 | 1249 | 33 | US-60-385-568-357 | Sequence 357, App |
| 39 | 100 | 12.8 | 1266 | 27 | US-10-179-131-9317 | Sequence 9317, Ap |
| 40 | 98.5 | 12.6 | 850 | 20 | US-09-614-150-24084 | Sequence 24084, A |
| 41 | 98.5 | 12.6 | 850 | 20 | US-09-614-150A-24084 | Sequence 24084, A |
| 42 | 98.5 | 12.6 | 850 | 33 | US-60-151-637-24179 | Sequence 24179, A |
| 43 | 98.5 | 12.6 | 850 | 33 | US-60-191-681-19019 | Sequence 19019, A |
| 44 | 98.5 | 12.6 | 1028 | 20 | US-09-614-150-14916 | Sequence 14916, A |
| 45 | 98.5 | 12.6 | 1028 | 20 | US-09-614-150A-14916 | Sequence 14916, A |

ALIGNMENTS

RESULT 1
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043 406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

```

US-09-543-407-12
Query Match      100.0%; Score 779; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.9e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
QY 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151

RESULT 2
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      89.3%; Score 696; DB 19; Length 151;
Best Local Similarity 87.6%; Pred. No. 2e-65;
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
QY 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 117
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
DB 118 -----YDQLVTRVVTHEMAHASVMVQVGFNNATANQY 151

RESULT 3
US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match      89.2%; Score 695; DB 19; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.5e-65;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAQ 60
QY 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SPARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNYDQVTRVVTHEMAHANNATANQY 151

RESULT 4
US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

Query Match      88.6%; Score 690; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 8.6e-65;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

```



```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNPALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 5
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24
Query Match 84.1%; Score 655; DB 19; Length 151;
Best Local Similarity 79.8%; Pred. No. 4.6e-61;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 109
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 110 -----NYDQLVTRVVTHEMAHANNATANQY 151

RESULT 6
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26
Query Match 78.8%; Score 614; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.1e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 LVTRVVTHEMAHAGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18
Query Match 78.7%; Score 613; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.4e-56;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 20
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match          78.4%; Score 611; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.2e-56;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKYDQLVTRVVTHEMAHAGQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNVQDLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNVQTSDDSSVVMVROVGFNNATANQY 151

RESULT 9
US-09-543-407-31
/ Sequence 31, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 31
/ LENGTH: 131
/ TYPE: PRT
/ ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match          78.0%; Score 608; DB 19; Length 131;
Best Local Similarity 88.5%; Pred. No. 3.9e-56;
Matches 116; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60

QY 81 VQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNVQDLVTRVVTHEM 140
Db 61 VQGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNVQTSDDSSVVMVROV 120

QY 141 AHANNATANQY 151
Db 121 GFGNNATANQY 131
```

```
RESULT 10
US-09-543-407-28
/ Sequence 28, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 28
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match          77.7%; Score 605; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 9.8e-56;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNVQDLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNVQTSDDSSVVMVROVGFNNATANQY 151

RESULT 11
US-09-543-407-22
/ Sequence 22, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match          77.5%; Score 604; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
Qy 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNDQASDSSVMVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30

Query Match 77.4%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Qy 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNDQASDSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16
Query Match 74.5%; Score 580; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 4.5e-53;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVDLVTRVVTHEMAHANNATANQY 151
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Qy 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNDQASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-55

Query Match 71.9%; Score 560; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.5e-51;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 22 VVPQGGGNGHNGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADV 81

Db 1 VVPQMGGGNHGNGSGPDSLTSLIYQGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVN 127
Db 61 GQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVN 106

RESULT 15

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 66.8%; Score 520; DB 19; Length 151;
Best Local Similarity 66.9%; Pred. No. 1,1e-46;
Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGHNGGNSGPNSELNLYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 TDARNSDLTITQGGGNGADVQGGSDSDSIDLTQRFNGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151
Db 121 GNCAAVDQTAASNSVNVTVQVFGNNATAHQY 151

Search completed: August 2, 2004, 15:26:42
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....VTRVVTHEMAHANNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_AA_New:*
1: /cgn2.6/prodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2.6/prodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2.6/prodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2.6/prodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2.6/prodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2.6/prodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2.6/prodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|----------------------|
| 1 | 517 | 66.4 | 151 | 5 | US-09-741-873C-4 |
| 2 | 439 | 56.4 | 131 | 5 | US-09-741-873C-2 |
| 3 | 97.5 | 12.5 | 258 | 6 | US-10-425-115-300390 |
| 4 | 97.5 | 12.5 | 295 | 6 | US-10-425-115-312488 |
| 5 | 94.5 | 12.1 | 719 | 5 | US-09-248-796A-17559 |
| 6 | 92 | 11.8 | 299 | 6 | US-10-170-205E-35751 |
| 7 | 92 | 11.8 | 1203 | 6 | US-10-170-205E-741 |
| 8 | 92 | 11.8 | 1327 | 1 | PCT-US04-02338-49 |
| 9 | 91.5 | 11.7 | 256 | 6 | US-10-425-115-301334 |
| 10 | 91.5 | 11.7 | 443 | 6 | US-10-100-683-7608 |
| 11 | 91.5 | 11.7 | 841 | 7 | US-60-565-632-7906 |
| 12 | 91.5 | 11.7 | 841 | 7 | US-60-579-062-7906 |
| 13 | 90.5 | 11.6 | 234 | 6 | US-10-767-701-45603 |
| 14 | 90 | 11.6 | 586 | 1 | PCT-US03-24982A-317 |
| 15 | 88 | 11.3 | 573 | 7 | US-60-565-632-7907 |
| 16 | 88 | 11.3 | 573 | 7 | US-60-579-062-7907 |
| 17 | 87.5 | 11.2 | 179 | 6 | US-10-425-115-346132 |
| 18 | 87.5 | 11.2 | 892 | 5 | US-09-952-267B-5 |
| 19 | 87.5 | 11.2 | 892 | 6 | US-10-872-768-5 |
| 20 | 87.5 | 11.2 | 892 | 6 | US-10-872-768-5 |
| 21 | 86 | 11.0 | 412 | 7 | US-60-565-632-7905 |
| 22 | 86 | 11.0 | 412 | 7 | US-60-579-062-7905 |
| 23 | 86 | 11.0 | 520 | 6 | US-10-479-638-21 |
| 24 | 85 | 10.9 | 956 | 6 | US-10-093-037A-63 |
| 25 | 84.5 | 10.8 | 251 | 6 | US-10-854-439-256 |
| 26 | 83.5 | 10.7 | 251 | 6 | US-10-854-439-249 |

27 82.5 10.6 873 5 US-09-952-267B-13 Sequence 13, Appl
28 82.5 10.6 873 6 US-10-872-768-13 Sequence 13, Appl
29 82.5 10.6 873 6 US-10-872-768-13 Sequence 13, Appl
30 82 10.5 193 6 US-10-425-115-354240 Sequence 234240, A
31 82 10.5 386 6 US-10-100-683-10326 Sequence 10326, A
32 82 10.5 386 6 US-10-100-683-10327 Sequence 10327, A
33 82 10.5 386 6 US-10-798-512-100 Sequence 100, Appl
34 82 10.5 555 1 PCT-US04-13211-3 Sequence 3, Appl
35 82 10.5 555 6 US-10-835-904-3 Sequence 3, Appl
36 82 10.5 1627 6 US-10-170-205E-16659 Sequence 16659, A
37 82 10.5 1305 1 PCT-US04-09388-9 Sequence 9, Appl
38 81.5 10.5 376 6 US-10-491-733-2 Sequence 2, Appl
39 81.5 10.5 508 6 US-10-425-115-285216 Sequence 285216, A
40 81.5 10.5 532 6 US-10-425-115-285214 Sequence 285214, A
41 81.5 10.5 1871 1 PCT-US03-02038-26 Sequence 26, Appl
42 81.5 10.5 4249 1 PCT-US04-04300-4 Sequence 4, Appl
43 81 10.4 131 7 US-60-565-632-11109 Sequence 11109, A
44 81 10.4 131 7 US-60-579-062-11109 Sequence 11109, A
45 81 10.4 201 6 US-10-425-115-309662 Sequence 309662, A

ALIGNMENTS

RESULT 1

US-09-741-873C-4

; Sequence 4: Application US/09741873C

; GENERAL INFORMATION:

; APPLICANT: Normark, Staffan

; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

; FILE REFERENCE: 012889-084

; CURRENT APPLICATION NUMBER: US/09/741,873C

; CURRENT FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: SE 8801723-1

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: US 08/978,878

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 07/347,189

; PRIOR FILING DATE: 1989-05-04

; PRIOR APPLICATION NUMBER: US 07/789,437

; PRIOR FILING DATE: 1991-11-06

; PRIOR APPLICATION NUMBER: US 07/970,846

; PRIOR FILING DATE: 1992-11-03

; PRIOR APPLICATION NUMBER: US 08/187,865

; PRIOR FILING DATE: 1994-01-28

; PRIOR APPLICATION NUMBER: US 08/318,519

; PRIOR FILING DATE: 1994-10-05

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 4

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match 66.4%; Score 517; DB 5; Length 151;

Best Local Similarity 66.2%; Pred. No. 1.1e-36;

Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSTIYQVGSANAALQ 60

1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSTIYQVGSANAALQ 60

61 SPARSETTITOSGVNGADYGGADNSTIETLQNGFNNTATIDWNKXSDITVGOYGG 120

61 THARNSDITITQGGGNGADVGGSDSSIDLTQRGFGNSATLDQWNGKSEMTVKQFGG 120

121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

121 GNAAVDQTASNSVNVTVGVGNNATANQY 151

```
RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012899-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8501723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRP
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match      56.4%; Score 439; DB 5; Length 131;
Best Local Similarity 62.6%; Pred. No. 3.9e-30;
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVVPQGGGNNHGGNSGSPDSTLSIYQGSANAALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQGGGNNHGGNSGSPDSTLSIYQGSANAALQSDARKSETTITQSGYNGAD 60

QY 81 VGGGADNSTIELTQGFNRNATIDOWNAKNSDITVQYGGNNAALVNDQLVTRVTHEM 140
Db 61 VGGGSDSSIDITQGFNSALTDQNGKNSMTVKQFGGNGRAVDQTASNSVNTQV 120

QY 141 AHANNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 3
US-10-425-115-300390
; Sequence 300390, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300390
; LENGTH: 258
; TYPE: PRP
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(258)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37025C.1.pap
```

US-10-425-115-300390

```
Query Match      12.5%; Score 97.5; DB 6; Length 258;
Best Local Similarity 26.6%; Pred. No. 0.86;
Matches 34; Conservative 18; Mismatches 53; Indels 23; Gaps 5;

QY 17 SALAGVVPQCGGNNHGGNSGSPDSTLSIYQGSANAALQSDARKSETTITQSG-- 74
Db 78 SSIAG-----GGGGQGGGGTNGGSGSGGGSGYVGSSTAAASGPGSGNYADAEGKGAG 132

QY 75 -----YNGA--DVGGADNSTIEL--TQGFNRNATIDOWNAKNSDITVGOYGG 120
Db 133 GGMGGGANGAYGSGAGGVKGGEVSGVALAPNSGYNGAADATGGGSG--AGGGHGG 190

QY 121 NNAALVNY 128
Db 191 GAAGAPSY 198
```

RESULT 4

```
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRP
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pap
US-10-425-115-312468
```

```
Query Match      12.5%; Score 97.5; DB 6; Length 295;
Best Local Similarity 29.7%; Pred. No. 1;
Matches 38; Conservative 15; Mismatches 50; Indels 25; Gaps 4;

QY 26 WGGGNNHGGNSGSPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGGGA 85
Db 138 YGGGYSSGGYSSG-GYAANGYGVSGSGNYSNASGGYSGS-----DYGNGAASGGYA 192

QY 86 DNSTIELTQGFNRNATIDOWNAKNSDITVQYGGNNAAL 125
Db 193 NNLSSGNSGRYNTIGSSDGTGGYNSYPNPGAGNTYNTGGSSSSGGTGFEGGGGFGG 252

QY 126 VNYDQLVT 133
Db 253 NNFAGNVT 260
```

RESULT 5

```
US-09-248-796A-17559
; Sequence 17559, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
```

; SEQ ID NO 17559
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17559

Query Match 12.1%; Score 94.5; DB 5; Length 719;
Best Local Similarity 25.8%; Pred. No. 4.9;
Matches 31; Conservative 24; Mismatches 42; Indels 23; Gaps 4;

QY 30 GNHNGGNSGPDSTLSIYQGSANAALOSDARKSETTITQSGYNGADVGQGDNST 89
DB 519 GNNNGSGNSTTNNNNYNN-----KSISKQNE-----IDGDLNPTSTINN 562
QY 90 IELTQNGFRNATIDQNAKNSDITVGOYGGNNAALVNYDOLVTRVVTHEMAHANNTAN 149
DB 563 TGLTNN-----NNSKSPAKSKKSNFD-----NNSNALNLDKSLKINTEITNITSETSN 615

RESULT 6
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

Query Match 11.8%; Score 92; DB 6; Length 299;
Best Local Similarity 30.4%; Pred. No. 3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNATIDQNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 7
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

Query Match 11.8%; Score 92; DB 6; Length 1203;
Best Local Similarity 30.4%; Pred. No. 14;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64

DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNATIDQNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 8
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPCAXS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; FILE OF INVENTION: USE
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,850
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match 11.8%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 16;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNSPSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQGDNSTIELTQNG--FRNATIDQNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 9
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pcp
US-10-425-115-301334

Query Match 11.7%; Score 91.5; DB 6; Length 256;
Best Local Similarity 26.6%; Pred. No. 2.8;

```
Matches 33; Conservative 16; Mismatches 52; Indels 23; Gaps 5;
QY 17 SALAGVVPWGGGNGHNGSSGPDSTLSIYQGSANAALQSDA----RKSETTTIQ 72
Db 79 SSVAG-----GGGGGGGGGTTNGSGSGSGSGTSTTAASGSPSGNYANAGKAG 133
QY 73 SGVGNAGD-----VQGGADNSTIEL--TQGNFRNNATIDQNNKNSDITVQYGG 120
Db 134 GGMGGGADGAYSGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 191
QY 121 NNA 124
Db 192 GAG 195

RESULT 10
US-10-100-683-7608
; Sequence 7608, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7608
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-100-683-7608

Query Match 11.7%; Score 91.5; DB 6; Length 443;
Best Local Similarity 26.0%; Pred. No. 5.1;
Matches 33; Conservative 18; Mismatches 43; Indels 33; Gaps 6;
QY 21 GVVPQ---WGGGNGHNGSSGPDSTLSIYQ--YGSANAALQSDARKSETTTQSGY 75
Db 185 GMPQAPAGQGG--NGGPPNFTNTQGAAPGYSVRA-----SNQNEGCTNPPPSGS 237
QY 76 GNGAD---VQGGADNSTIELTQGNFRNNATI-----DQNNKNSDIT 114
Db 238 GGSNSNGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSS 297
QY 115 VQYQGN 121
Db 298 SGNHGS 304

RESULT 11
US-60-565-632-7906
; Sequence 7906, Application US/60565632
```

```
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Muryikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7906

Query Match 11.7%; Score 91.5; DB 7; Length 841;
Best Local Similarity 32.6%; Pred. No. 11;
Matches 44; Conservative 12; Mismatches 48; Indels 31; Gaps 9;
QY 30 GNHGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTTQSGYNGAD 80
Db 269 GNENGGAENNAADACTDAAQ--GSTNEAENNAADVQNDAAQENGAENGAENSGAD 327
QY 81 VQGGADN--STIELTON-----GFRNNATIDQNNKNSDITVQYQGNNAALVNYDQVTR 134
Db 328 AAQCTDNGAAAGTGNADPAQNDNGAA-----AENSGNENGTAAGNNA---NPD----- 374
QY 135 VVTHEMAHAN--NATA 148
Db 375 -VQNDAAQVNDNGTA 388

RESULT 12
US-60-579-062-7906
; Sequence 7906, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K
; APPLICANT: Larosa, Thomas J
; APPLICANT: Lu, Maolong
; APPLICANT: Muryikwa, Tichifa R. I.
; APPLICANT: Roberts, James K
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7906
```



```

Query Match      11.7%; Score 91.5; DB 7; Length 841;
Best Local Similarity 32.6%; Pred. No. 11;
Matches 44; Conservative 12; Mismatches 48; Indels 31; Gaps 9;

QY 30 GNHGGG--NSSGPDSTLSIQYGSANAALA-----LQSDARKSETTITQSGYGNAD 80
DB 269 GNENGTGAENNAADAQTDAAQ--GSTNEAENNAADVQNDAAQAENAGAAENSGNAD 327
QY 81 VQGGADN--STIELTQN-----GFRNNATIDWNAKNSDITVQYQYGNNAALVNYDQLVTR 134
DB 328 AAGCTDNGAAAGTGNADPAQNDNGAA-----AENSGNENGTAAGNNA---NPD----- 374
QY 135 VVTHEMAHAN-NATA 148
DB 375 -VQNDAAQVNDNGTA 388

RESULT 13
US-10-767-701-45603
; Sequence 45603, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45603
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(234)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C62834_1.pep
US-10-767-701-45603

Query Match      11.6%; Score 90.5; DB 6; Length 234;
Best Local Similarity 24.2%; Pred. No. 3;
Matches 39; Conservative 12; Mismatches 45; Indels 65; Gaps 6;

QY 1 MKLLKVAFAAATVW--SGSALAGVVPQW---GGGNGNHGGGN-----SSGPDSTL 45
DB 1 MATTKLALCFVLLIGIGANARVARVVSAGGGGGGGGGRGWRGASRWGSGSGGC 60
QY 46 STY-----QYGSANAALALQSDARKSETTITQSGYGNADYVG 82
DB 61 GXYGAGSGGXAYAQGGGGGGGGQYGGSG-----SGYSGSGYV 102
QY 83 QGADNSTIELTQNGFRNNATIDWNAKNSDITVGOYGNNA 123
DB 103 QAGSGS-----NGGAYAQGAQGGGGGGGGQYGGSGS 134

RESULT 14
PCT-US03-24982A-317
; Sequence 317, Application PC/TUS0324982A
; GENERAL INFORMATION:
; APPLICANT: Syngenta Participations AG
; APPLICANT: Stam, Lynn
; APPLICANT: Kamdar, Kim
; APPLICANT: Spana, Eric
; APPLICANT: Bachmann, Jane
; TITLE OF INVENTION: Nucleic Acid Sequences from Drosophila Melanogaster that E
; TITLE OF INVENTION: Essential for Viability and Uses Thereof
; FILE REFERENCE: 7013WOPCT
; CURRENT APPLICATION NUMBER: PCT/US03/24982A
; CURRENT FILING DATE: 2003-08-08

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTTHMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 695 | 89.2 | 151 | JC6039 | fimbrin protein ag |
| 2 | 695 | 89.2 | 151 | A10635 | major curlin chain |
| 3 | 520 | 66.8 | 151 | S70788 | curlin protein csg |
| 4 | 501.5 | 64.4 | 152 | D90806 | curlin major subun |
| 5 | 501.5 | 64.4 | 152 | H85665 | hypothetical prote |
| 6 | 106 | 13.6 | 1748 | S42136 | cnjB protein - Tet |
| 7 | 104 | 13.4 | 2174 | E95965 | hypothetical glyci |
| 8 | 101.5 | 13.0 | 151 | AH0635 | nucleation compone |
| 9 | 100.5 | 12.9 | 151 | JC6040 | fimbrin protein ag |
| 10 | 98.5 | 12.6 | 1028 | A56038 | DNA-binding protei |
| 11 | 98.5 | 12.6 | 1213 | S16356 | ovo protein - frui |
| 12 | 96 | 12.3 | 145 | AD3143 | conserved hypothet |
| 13 | 96 | 12.3 | 145 | H98144 | hypothetical prote |
| 14 | 94.5 | 12.1 | 678 | A70762 | probable PPE prote |
| 15 | 93.5 | 12.0 | 1238 | AH0038 | probable PPE prote |
| 16 | 93 | 11.9 | 573 | C86266 | F3f19.21 protein - |
| 17 | 91.5 | 11.7 | 151 | S70787 | curlin nucleator p |
| 18 | 91.5 | 11.7 | 151 | C90806 | minor curlin subun |
| 19 | 91.5 | 11.7 | 151 | C85665 | curlin minor chain |
| 20 | 91.5 | 11.7 | 256 | T03371 | glycine-rich prote |
| 21 | 91 | 11.7 | 770 | T51034 | related to C2H2 z1 |
| 22 | 91 | 11.7 | 1250 | D91018 | hypothetical prote |
| 23 | 91 | 11.7 | 1250 | F85862 | hypothetical prote |
| 24 | 90.5 | 11.6 | 745 | E64559 | outer membrane pro |
| 25 | 90 | 11.6 | 409 | T20847 | hypothetical prote |
| 26 | 90 | 11.6 | 573 | S35327 | protein kinase sgg |
| 27 | 90 | 11.6 | 582 | F70675 | probable PPE prote |
| 28 | 89.5 | 11.5 | 262 | S00275 | tail fiber protein |
| 29 | 89.5 | 11.5 | 343 | T05221 | hypothetical prote |

```

30      89      11.4      347      2      B39112      merozoite 45K surf
31      89      11.4      3300      2      D70575      probable PPE prote
32      88.5      11.4      407      2      T21956      hypothetical prote
33      88.5      11.4      447      2      G84687      probable disease r
34      88.5      11.4      552      2      D70604      probable PPE prote
35      88      11.3      301      2      B84533      hypothetical prote
36      88      11.3      354      2      B70663      probable PPE prote
37      88      11.3      434      2      E70768      hypothetical glyci
38      87.5      11.2      615      2      E70663      probable PPE prote
39      87.5      11.2      940      2      D89723      protein F39D8.1b f
40      87.5      11.2      945      2      T21998      hypothetical prote
41      87      11.2      331      2      S21409      class 3 outer memb
42      87      11.2      440      2      AD1539      probable sugar ABC
43      87      11.2      645      2      F70825      probable PPE prote
44      86.5      11.1      1635      2      AI8452      hemolysin (importe
45      86.5      11.1      3190      2      T13828      CREB-binding prote

```

ALIGNMENTS

RESULT 1

JC6039 fimbrin protein agfA precursor - Salmonella enteritidis

C;Species: Salmonella enteritidis

C;Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999

C;Accession: JC6039; PC6015; J44898

R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.

J. Bacteriol. 179, 562-667, 1996

A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.

A;Reference number: JC6039; MUID:96146512; PMID:8550497

A;Accession: JC6039

A;Molecule type: DNA

A;Residues: 1-151 <COL>

A;Cross-references: GB:U43280; NID:gl184712; PIDN:ARC43599.1; PID:gl184714

A;Accession: PC6015

A;Molecule type: protein

A;Residues: 21-52 <CO2>

A;Experimental source: strain 27655-3b

A;Note: the authors translated the codon ACG for residue 44 as Ile

R;Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A;Title: Purification and characterization of thin, aggregative fimbrinae from Salmonell

A;Reference number: A44898; MUID:91310586; PMID:1677357

A;Contents: 27655

A;Accession: A44898

A;Status: preliminary

A;Molecule type: protein

A;Residues: 21-33 <CO3>

A;Note: sequence extracted from NCBI backbone (NCBIP:45936)

C;Genetics:

A;Gene: agfA

C;Function:

A;Description: major component of thin aggregative fimbrinae

A;Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator

C;Keywords: fimbrina

F;1-20/Domain: signal sequence #status predicted <SIG>

F;21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match

Best Local Similarity 89.2%; Score 695; DB 2; Length 151;

Mismatches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAIQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAIQ 60

QY 61 SPARKSETTITOSGYNGADVCGQADNSTIETQTQGFNNATIDOWNAKNSDITVGYCGG 120

Db 61 SPARKSETTITOSGYNGADVCGQADNSTIETQTQGFNNATIDOWNAKNSDITVGYCGG 120

QY 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151

Db 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151

Db 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

A:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Mouton, S.; O'Garra, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 89.2%; Score 695; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 1.8e-50;

Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Db 61 SDARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Qy 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151

Db 121 NNAALVNYDQTSADSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csga precursor - Escherichia coli (strain K-12)

N:Alternate names: csga protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csq operons is required for production of fibronectin- and collagen-binding proteins by Escherichia coli

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA074126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csqA

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RQRSGMWLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:G290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csGA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between csGA and csGB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of csGA, csGB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 66.8%; Score 520; DB 2; Length 151;

Best Local Similarity 66.3%; Pred. No. 4.7e-36;

Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGGNGSGGPDSTLSIYQGGNSALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Db 61 TDARNSDLTITQGGGNGADVGQGDSDSIDITQRFNGSATLDQWNGKNSKEMTKVQFGG 120

Qy 121 NNAALVNYDQLVTRVVTTHMAHANNATANQY 151

Db 121 GNGAAVDQTSASNVVTVQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit csGA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs1420

Query Match 64.4%; Score 501.5; DB 2; Length 152;

Best Local Similarity 65.8%; Pred. No. 1.6e-34;

Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGNGGNGSGGPDSTLSIYQGSANAALAL 59

Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGGNGSGGPDSTLSIYQGGNSALAL 60

Qy 60 QSDARKSETTITQSGYNGADVGQGDADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 119

```
Db 61 QADAENSLDTITQHGGNGADVGQGGDDSSIDLQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 120 GNNAAALVNYDQLVTRVTHMAHANNATANOY 151
Db 121 GNGRAVDQATASNTVNTVQVFGNNATAHQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Finan, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AR005174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UWGP:216
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match 64.4%; Score 501.5; DB 2; Length 152;
Best Local Similarity 65.8%; Pred. No. 1.6e-34;
Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAALVVGSGALAGVPOW-QGGGNHNGGNSGPDSTLSIYQVGSANAALAL 59
Db 1 MKLLKVAFAALVVGSGALAGVPOYGGGGNGHGGGNSGPNSELNIYQVGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVGQGGADNSTIELQNGFRNNATIDQWNAKNSDITVQYQ 119
Db 61 QADARNSLDTITQHGGNGADVGQGGDDSSIDLQRFNGSATLDQWNGKDSHMTVKQFG 120

QY 120 GNNAAALVNYDQLVTRVTHMAHANNATANOY 151
Db 121 GNGRAVDQATASNTVNTVQVFGNNATAHQY 152

RESULT 6
cniB protein - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C:Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
Submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
A:Residues: 1-1748 <TAY>
A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:86189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250, 'I', 252-255, 'N', 257-773 <MAR>
```

```
A:Cross-references: EMBL:X06462
C:Genetics:
A:Gene: cniB
A:Genetic code: SGCS
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C:Keywords: zinc finger
F:1164-1450/Region: glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 13.6%; Score 106; DB 2; Length 1748;
Best Local Similarity 32.1%; Pred. No. 1;
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;

QY 25 QWGGGNHNGG---GNSSGPDSTLSIYQVGSANAALALQSDARKSETTIT---QSGYGN 77
Db 1640 QFGGGNSNGGQSGWTSSGSDWN-----CQSNVQESTTSSGGWGGSGSGN 1685

QY 78 GADVGGGADNSTIELQNGFRNNATIDQWNAKNSDITVQG--VGGNAA 124
Db 1686 QTGGGWSNDN-----QQQENNTGGGGWSSNSQTNNESSWGSNNQA 1729

RESULT 7
E95965
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) magi
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95965
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95965
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2174 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:g15140875; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
Peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb21548
A:Genome: plasmid

Query Match 13.4%; Score 104; DB 2; Length 2174;
Best Local Similarity 27.1%; Pred. No. 1.9;
Matches 35; Conservative 16; Mismatches 44; Indels 34; Gaps 6;

QY 11 AIWVSGSALAGVVPQ--WGCGGNHNGGNSGPDSTLSIYQVGSANAALALQSDARKSET 68
Db 693 AIATAGAGAVGILAQISIGGGGN--GGNATGGDAGFGSFGQGGGGG----- 737

QY 69 TITQSGYNGADVG-----QGADNSTI--ELTQNGFRNNATIDQWNAK---NSDITV 115
Db 738 ----GGYANTANVGKGLTLTTQGSRAAGIVAQSVGGGGGTGTTASSYSAGIGFTASVAV 793

QY 116 GOYVGNAA 124
Db 794 GGTGGNGGA 802
```

```
RESULT 8
AH0635
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Notes: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
  T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
  S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Barry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0635
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl6502314; GSPDB:GN00176
C:Genetics:
A:Gene: STY1180

Query Match 13.0%; Score 101.5; DB 2; Length 151;
Best Local Similarity 30.2%; Pred. No. 0.15;
Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;

QY 47 IVYGSANAALQSDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDOW 106
DB 49 IGQVGTDSNA-RVROGSKLLSVISOEGNNRKAQVDQAGNYNFAYIEQTGNANDASISQS 107

QY 107 NAKNSDITVQYGGNNAALVNDQLVTRVVTHEMAH 142
DB 108 AYGNRAITQKSGNKANITQYGTQKTAVVQKQSH 143

RESULT 9
JC6040
fimbriin protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A>Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43598.1; PID:gl1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
A:Function:
A>Description: minor component of thin aggregative fimbriae
A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbriin protein agfB #status predicted <MAT>

Query Match 12.9%; Score 100.5; DB 2; Length 151;
Best Local Similarity 30.2%; Pred. No. 0.18;
Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;

QY 47 IVYGSANAALQSDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDOW 106
DB 49 IGQVGTDSNA-RVROGSKLLSVISOEGNNRKAQVDQAGNYNFAYIEQTGNANDASISQS 107

QY 107 NAKNSDITVQYGGNNAALVNDQLVTRVVTHEMAH 142
DB 108 AYGNRAITQKSGNKANITQYGTQKTAVVQKQSH 143
```

```
RESULT 10
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C:Accession: A56038
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell Biol. 14, 6809-6818, 1994
A>Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster
A:Reference number: A56038; MUID:95021209; PMID:7935398
A:Accession: A56038
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1028 <GAR>
A:Cross-references: GB:U11383; NID:G520526; PIDN:AAB60216.1; PID:G520527
C:Genetics:
A:Gene: ovo
A:Cross-references: FlyBase:FBgn0003028

Query Match 12.6%; Score 98.5; DB 2; Length 1028;
Best Local Similarity 24.5%; Pred. No. 2.3;
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;

QY 3 LLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANAALALQSD 62
DB 59 LQNAAAAYIMSAGS-----GGCTGCGGGGASGPGGSPANSNGGGGG----- 104

QY 63 ARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDOWNAKNSDI----- 113
DB 105 -----CGNGYINGGVG-GENNS--LDGNNLLNFASVSNYSNKSFKFNHHHHQH 152

QY 114 -----TVQG-----YCGNNAAL-----VNYDQ-----LVTRVVTHEMAH 143
DB 153 NNNNNNGGQTSMMVGHFYGNGPSAYGIILKQDPDIEYDEAKIDICTFAQNIQATMGSS 212

QY 144 NNATANQY 151
DB 213 GQFNASAY 220

RESULT 11
S16356
ovo protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C:Accession: S16356
R:Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2259-2266, 1991
A>Title: The ovo gene of Drosophila encodes a zinc finger protein required for female g
A:Reference number: S16356; MUID:91293102; PMID:1712294
A:Accession: S16356
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1213 <MEV>
A:Cross-references: EMBL:X59772
C:Genetics:
A:Gene: FlyBase:ovo
A:Cross-references: FlyBase:FBgn0003028
A:Introns: 931/3; 1152/3

Query Match 12.6%; Score 98.5; DB 2; Length 1213;
Best Local Similarity 24.5%; Pred. No. 2.8;
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;

QY 3 LLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGSSGPDSTLSIYQGSANAALALQSD 62
DB 422 LQNAAAAYIMSAGS-----GGCTGCGGGGASGPGGSPANSNGGGGG----- 467

QY 63 ARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDOWNAKNSDI----- 113
DB 468 -----CGNGYINGGVG-GENNS--LDGNNLLNFASVSNYSNKSFKFNHHHHQH 515
```

```
QY 114 -----TVGQ-----YGGNAAAL-----VNYDQ-----LVTRVVTHEMAHA 143
D 516 NNNNNNGGQSMGHPPYGNPSPAYGILKDEPDIEYDEAKIDIGTFAQNIQATWGSS 575
QY 144 NNATANQY 151
D 576 GQFNASAY 583

RESULT 12
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3143
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AR008689; PIDN:AAL45562.1; PID:g17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;
Best Local Similarity 27.1%; Pred. No. 0.41;
Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;

QY 3 LLKVAAPAAIVVSGSALAGVVP-----OWG-----GGNHNGGN-----SSGPDST 44
D 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYGYNSAGGAEQGVGNRIYQNGYNR 60
QY 45 LSIYQYSANAALQSDARKSETTITQSGYGNAGDVQGGADNSTIELTQNGFRNNATID 104
D 61 IVGHQYGRHNLS-AVGQEGHDNYGTTQNGRNVAGIQFGSNHTTILTDGNGNIAAGV 119

QY 105 QWNAKNSDITVGQYGGNAA 124
D 120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 13
H98144
Hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98144
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98144
A>Status: preliminary
A:Cross-references: GB:AR007870; PIDN:AAK88682.1; PID:g15158413; GSPDB:GN00170
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89171.1; PID:g15978410; GSPDB:GN00175
C:Genetics:
A:Gene: AGR_L_228
A:Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;
Best Local Similarity 27.1%; Pred. No. 0.41;
Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;
```

```
QY 3 LLKVAAPAAIVVSGSALAGVVP-----OWG-----GGNHNGGN-----SSGPDST 44
D 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYGYNSAGGAEQGVGNRIYQNGYNR 60
QY 45 LSIYQYSANAALQSDARKSETTITQSGYGNAGDVQGGADNSTIELTQNGFRNNATID 104
D 61 IVGHQYGRHNLS-AVGQEGHDNYGTTQNGRNVAGIQFGSNHTTILTDGNGNIAAGV 119

QY 105 QWNAKNSDITVGQYGGNAA 124
D 120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 14
A70762
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: A70762
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scars, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70762
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-678 <COL>
A:Cross-references: GB:Z74020; GB:AL123456; NID:g3261584; PIDN:CAA98335.1; PID:e1300075
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 12.1%; Score 94.5; DB 2; Length 678;
Best Local Similarity 25.7%; Pred. No. 3.1;
Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;

QY 15 SGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYSANAALQ--SDARKSETTITQ 72
D 384 SSGSGLG---FGNSGNGIGFFNSG--NNNIGMNSGNGVGALSVEFGSAAERS----- 432
QY 73 SGYGNAGDVQGGADNS-----TIELTQNGFRNNATIDQ--WNAKNSDITVGQYGGNN 122
D 433 SGFGSGELSTGIGNSGQLSTGWFNSATTSTGWFNSGTTTGTGFWNSGTTTGTGNSGGN- 491
QY 123 AALVNYDQIVTRVVTHEM-----AHANNATAN 149
D 492 -----LVTGSMGLFNSGHTNTGSPN 511

RESULT 15
AH0038
probable exported protein YPO0309 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AH0038
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0038
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1238 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89171.1; PID:g15978410; GSPDB:GN00175
C:Genetics:
A:Gene: YPO0309
```

```

Query Match      12.0%; Score 93.5; DB 2; Length 1238;
Best Local Similarity 25.1%; Pred. No. 7.3;
Matches 45; Conservative 19; Mismatches 48; Indels 67; Gaps 9;

QY 7 AAFALVSGSALAGVVPWGCGG---NHNGGN-----SSG-----P 41
Db 409 SAFSAITATGHLTA---EWQGAMLQTHSSLGDATLHFNDITAMSSGISLINEANQG 464
QY 42 DSTLSIYQYGSANAA-----LALQSDARKSETTITQSGYGN 77
Db 465 TSTADITVIGQINVSHGEGITLNALTTDGRTLVNVVDVNNIASEYDAIRLYNNYNDNYAT 524
QY 78 GADVGGQADN--STIEL-----TQNGFRNNATIDQWNAKNSDITVG--QYGGNNAAL 125
Db 525 GVDDGTGADNGTSTDLITRGALVSGQGYGINI---ETNTADTYVTVGGLVHGGNGTAI 580

```

Search completed: August 2, 2004, 14:56:21
Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTHMAFANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 695 | 89.2 | 151 | 1 CSGA_SALTY | P52225 salmonella |
| 2 | 520 | 66.8 | 151 | 1 CSGA_ECOLI | P28307 escherichia |
| 3 | 501.5 | 64.4 | 152 | 1 CSGA_ECO57 | Q93u24 escherichia |
| 4 | 101.5 | 13.0 | 151 | 1 CSGB_SALTY | Q827m3 salmonella |
| 5 | 100.5 | 12.9 | 151 | 1 CSGB_SALTY | P52226 salmonella |
| 6 | 98.5 | 12.6 | 1028 | 1 OVO_DROME | P51521 drosophila |
| 7 | 97 | 12.5 | 1656 | 1 OMPB_RICJA | O96653 r outer mem |
| 8 | 94.5 | 12.1 | 678 | 1 YF48 MYCTU | Q10778 mycobacteri |
| 9 | 92 | 11.8 | 1327 | 1 TNK1_HUMAN | O85271 homo sapien |
| 10 | 91.5 | 11.7 | 151 | 1 CSGB_ECOLI | P39828 escherichia |
| 11 | 89.5 | 11.5 | 262 | 1 VQ38_EPT2 | P07875 bacterioph |
| 12 | 89 | 11.4 | 347 | 1 MGA2_PLAF2 | Q03646 plasmodium |
| 13 | 88 | 11.3 | 491 | 1 YK98_MYCTU | Q10707 mycobacteri |
| 14 | 87 | 11.2 | 331 | 1 OMB2_NEIMB | P30688 neisseria m |
| 15 | 86 | 11.0 | 1093 | 1 PER_DROWI | Q03297 drosophila |
| 16 | 85 | 10.9 | 1250 | 1 YFAL_ECOLI | P45508 escherichia |
| 17 | 84.5 | 10.8 | 3590 | 1 F9AB_BORPE | P12255 bordetella |
| 18 | 84 | 10.8 | 566 | 1 AMY_STRGR | P02070 streptomyce |
| 19 | 84 | 10.8 | 566 | 1 AMY_STRLM | P09794 streptomyce |
| 20 | 83.5 | 10.7 | 1067 | 1 SGB_DROME | P18431 drosophila |
| 21 | 83 | 10.7 | 590 | 1 GP63_LEIDO | P23223 leishmania |
| 22 | 83 | 10.7 | 599 | 1 GP63_LEIGH | P15706 leishmania |
| 23 | 83 | 10.7 | 602 | 1 GP63_LEIMA | P08148 leishmania |
| 24 | 83 | 10.7 | 646 | 1 GP63_LEIME | P43150 leishmania |
| 25 | 82.5 | 10.6 | 369 | 1 PST3_MYCAV | Q9kk89 mycobacteri |
| 26 | 82.5 | 10.6 | 794 | 1 YC84_MYCPN | P75493 mycoplasma |
| 27 | 82 | 10.5 | 1567 | 1 ICEN_XANCT | P18127 xanthomonas |
| 28 | 81.5 | 10.5 | 392 | 1 HME1_HUMAN | Q05925 homo sapien |
| 29 | 81.5 | 10.5 | 730 | 1 GLN3_YEAST | P18494 saccharomyc |
| 30 | 81 | 10.4 | 165 | 1 GRP1_ORYSA | P25074 oryza sativ |
| 31 | 81 | 10.4 | 385 | 1 PER_DRONE | P31686 drosophila |
| 32 | 81 | 10.4 | 485 | 1 YB47_MYCPN | P51339 mycoplasma |
| 33 | 81 | 10.4 | 959 | 1 N100_YEAST | Q02629 saccharomyc |

```

34      81      10.4      1460      1      PMPC_CHLMU
35      80.5      10.3      487      1      Y442_MYCTU
36      80.5      10.3      2003      1      YDBA_ECOLI
37      80      10.3      362      1      P35_MYCPE
38      80      10.3      594      1      SUH_DROME
39      80      10.3      955      1      FRU_DROME
40      80      10.3      1778      1      N189_SCHPO
41      79.5      10.2      485      1      Y136_TREPA
42      79.5      10.2      760      1      YB1L_ECOLI
43      79      10.1      572      1      FLAA_CAMCO
44      79      10.1      933      1      NPA3_HUMAN
45      79      10.1      1113      1      N116_YEAST

```

ALIGNMENTS

```

RESULT 1
CSGA_SALTY          STANDARD;          PRT;      151 AA.
ID      P52225;
AC      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Major curlin subunit precursor (Fimbrin SEF17).
GN      CSGA OR AGFA OR STM144 OR STY1181 OR TI1776.
OS      Salmonella typhimurium,
OS      Salmonella typhi, and
OS      Salmonella enteritidis.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Salmonella.
OX      NCBI_TaxID=602, 601, 592;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SPECIES=S.typhimurium; STRAIN=SR-11;
RX      MEDLINE=98117058; PubMed=9457880;
RA      Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT      "Curli fibers are highly conserved between Salmonella typhimurium and
RT      Escherichia coli with respect to operon structure and regulation.";
RL      J. Bacteriol. 180:722-731(1998).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX      MEDLINE=21534948; PubMed=11677609;
RA      McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA      Courteney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA      Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA      Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA      Waterston R., Wilson R.K.;
RT      "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT      LT2.";
RL      Nature 413:852-856(2001).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      SPECIES=S.typhi; STRAIN=CT18;
RX      MEDLINE=21534947; PubMed=11677608;
RA      Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA      Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA      Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA      Cronin A., Davis P., Davies K.M., Dowd L., White N., Farrar J.,
RA      Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA      Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA      Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA      Whitehead S., Barrall B.G.;
RT      "Complete genome sequence of a multiple drug resistant Salmonella
RT      enterica serovar Typhi CT18.";
RL      Nature 413:848-852(2001).
RN      [4]
RP      SEQUENCE FROM N.A.
RC      SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX      MEDLINE=22531367; PubMed=12644504;
RA      Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA      Burland V., Kocoyanni V., Schwartz D.C., Blattner F.R.;

```

```

Q9piy1 chlamydia m
P42611 mycobacteri
P33666 escherichia
Q50367 mycoplasma
P28159 drosophila
Q8in81 drosophila
Q9utk4 schizosacch
O83172 treponema p
P75780 escherichia
P27053 campylobact
O8ixf0 homo sapien
Q02630 saccharomyc

```

```

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL J. Bacteriol. 185:2330-2337(2003).
RN [5]
RC SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae";
RL J. Bacteriol. 178:662-667(1996).
RN [6]
RP SEQUENCE OF 21-151 FROM N.A.
RP SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis";
RL J. Bacteriol. 173:4773-4781(1991).
RC SEQUENCE OF 21-33.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis";
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL; AJ002301; CA05317.1; -
DR EMBL; A3008749; AAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; A3016840; AAC08399.1; -
DR EMBL; U43280; AAC43599.1; -
DR FIC; JC6039; JC6039.
DR StyGene; SG10608; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151
FT CHAIN 134 151
FT CONFLICT 134 151
FT REF. 6).
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;

Query Match 89.2%; Score 695; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.6e-51;
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYSANAALAQ 60
DB 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYSANAALAQ 60
QY 61 SDARKSETTITGSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITGSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNVQDLVTRVVTTHMANNATANQY 151
DB 121 NNAALVNVQDLVTRVVTTHMANNATANQY 151

```

```

RESULT 2
CSGA_ECOLI
ID _CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
RT repression of csgA, the subunit gene of fibronectin-binding curli in
RT Escherichia coli.";
RL Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=K12 / MC4100;
RX MEDLINE=96414458; PubMed=8817489;
RA Hammar M., Arngvist A., Sian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
RT K-12.";
RL Mol. Microbiol. 18:661-670(1995).
RN [3]
RP SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1232-1244(1997).
RN [4]
RP SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RP STRAIN=K12 / YNEM;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
RT fibronectin binding in Escherichia coli HB101.";
RL Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RP MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
RT Salmonella enteritidis";
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC

```

[illegible]


```
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69CDD5999F CRC64;
Query Match 12.5%; Score 97; DB 1; Length 1656;
Best Local Similarity 28.3%; Pred. No. 2.6;
Matches 43; Conservative 19; Mismatches 52; Indels 38; Gaps 7;
QY 6 VAFAAIVSGSALAGVWPWGCGGHHNGGSSGPDSTLSIYQYGSANAALQSDARK 65
Db 509 VLAAGAITDGSATI-----TGDICNGGG-----GAALQSITLANDATK 547
QY 66 SETTITQSG---YNGADVGQAGDNSTIETLQNGFRNNATIDQWNAKNSDITVVG--QVG 119
Db 548 ---TLILGGANIISANGTINFQANGTITKLTST--QNNIVD-----CDLAIATDQTG 596
QY 120 GNAALVNVYDQVTRVVVTHEMAHANNATANQY 151
Db 597 VWDASSLITNAQTLSITGTIGIIGANNTILGQF 628
RESULT 8
YF48_MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Hypochemical PPE-family protein Rv1548c/MT1599.
GN Rv1548C OR MT1599 OR MTC148.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; Z74020; CA98835.1; -.
CC EMBL; AE007026; AAK45866.1; ALT_INIT.
CC PIR; A70762; A70762.
CC TIGR; MT1599; -.
TubercuList; Rv1548c; -.
DR InterPro; IPR000030; Microbac PPE.
DR InterPro; IPR002989; Mycobac Pentapep.
DR Pfam; PF01469; Pentapeptide_2, 11.
DR Pfam; PF00823; PPE; 1.
KW Hypochemical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT CONFLICT 258 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
Query Match 12.1%; Score 94.5; DB 1; Length 678;
Best Local Similarity 25.7%; Pred. No. 1.5;
Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;
QY 15 SGSALAGVWPWGCGGHHNGGSSGPDSTLSIYQYGSANAALQ--SDARKSETTITQ 72
Db 384 SGSGNLG----FGNSGNGNIGFFNSG--NNNIGNSGNGVGLSVFSGSAERS----- 432
QY 73 SGYNGADVGQAGDNS-----TIELTQNGFRNNATIDQ--WNAKNSDITVQYGGNN 122
Db 433 SGFNSGELSTGIGNSQLSTGFWNSGATTSTGFWNSGTTTNGFWNSGTTTNGTGNSSGN- 491
QY 123 AALVNVYDQVTRVVVTHEM-----AHANNATAN 149
Db 492 -----LVTGSMGLFNSGHTNTGSFN 511
RESULT 9
TNK1_HUMAN STANDARD; PRT; 1327 AA.
AC O95271; O95272;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-
DE interacting ankyrin-related ADP-ribose polymerase).
GN TNKS OR TNKS1 OR TIN1 OR TINF1 OR PARPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Testis;
RX MEDLINE=99040105; PubMed=9822378;
RA Smith S., Giriat L., Schmitt A., de Lange T.;
RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
RL Science 282:1484-1487(1998).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99454782; PubMed=10523501;
RA Smith S., de Lange T.;
RT "Cell cycle dependent localization of the telomeric PAPP, tankyrase,
RT to nuclear pore complexes and centrosomes.";
RL J. Cell Sci. 112:3649-3656(1999).
RN [3]
RP FUNCTION, AND PHOSPHORYLATION.
RX MEDLINE=20556282; PubMed=10988299;
RA Chi N.-W., Lodish H.F.;
RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
RT substrate that interacts with IRAP in GLUT4 vesicles.";
RL J. Biol. Chem. 275:38437-38444(2000).
RN [4]
RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
RX MEDLINE=21602874; PubMed=1179745;
RA Cook B.D., Dynsek J.N., Chang W., Shostak G., Smith S.;
RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
RT at human telomeres.";
RL Mol. Cell. Biol. 22:332-342(2002).
CC -!- FUNCTION: May regulate vesicle trafficking and modulate the
CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PAPP
CC activity and can modify TRF1, and thereby contribute to the
```

CC regulation of telomere length
CC -!- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose} (N)-acceptor =
CC nicotinamide + {ADP-D-ribose} (N+1)-acceptor;
CC -!- SUBUNIT: Oligomerizes and associates with TNK52. Interacts with
CC the cytoplasmic domain of INPEP/Oase in SLC2A4/GLUT4-vesicles.
CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
CC -!- SUPRACELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
CC also found at nuclear pore complexes and around the pericentriolar
CC matrix of mitotic centrosomes. During interphase, a small fraction
CC of TNKS is found in the nucleus, associated with TRF1.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O95271-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95271-2; Sequence=VSP_004538, VSP_004539;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues
CC by MAPK kinases.
CC -!- PTM: ADP-ribosylated (-auto).
CC -!- SIMILARITY: Belongs to the PARP family.
CC -!- SIMILARITY: Contains 15 ANK repeats.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF082556; AAC79841.1; -;
CC EMBL; AF082557; AAC79842.1; -;
CC EMBL; AF082558; AAC79843.1; -;
CC EMBL; AF082559; AAC79844.1; -;
CC HSSP; Q00420; IAWC.
CC Genew; HGNC:11941; TNKS.
CC MIM; 603303; -;
CC GO; GO:0000781; C:chromosome, telomeric region; IDA.
CC GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
CC GO; GO:0005515; F:protein binding; IPI.
CC GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR001660; SAM.
CC Pfam; PF00023; ank; 19.
CC Pfam; PF00536; SAM; 1.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 17.
CC SMART; SM00454; SAM; 1.
CC PROSITE; PS50088; ANK REPEAT; 15.
CC PROSITE; PS50297; ANK REP REGION; 1.
CC PROSITE; PS50105; SAM DOMAIN; 1.
CC Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
CC Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
CC Phosphorylation; Alternative splicing.
CC FT REPEAT 215 247 ANK 1.
CC FT REPEAT 248 280 ANK 2.
CC FT REPEAT 281 313 ANK 3.
CC FT REPEAT 368 400 ANK 4.
CC FT REPEAT 401 433 ANK 5.
CC FT REPEAT 434 466 ANK 6.
CC FT REPEAT 521 556 ANK 7.
CC FT REPEAT 557 589 ANK 8.
CC FT REPEAT 590 622 ANK 9.
CC FT REPEAT 683 715 ANK 10.
CC FT REPEAT 716 748 ANK 11.
CC FT REPEAT 749 781 ANK 12.
CC FT REPEAT 836 868 ANK 13.
CC FT REPEAT 869 901 ANK 14.
CC FT REPEAT 902 934 ANK 15.

FT DOMAIN 1030 1089 SAM.
FT DOMAIN 1176 1327 PARP.
FT DOMAIN 9 14 POLY-HIS.
FT DOMAIN 27 34 POLY-PRO.
FT DOMAIN 128 134 POLY-SER.
FT DOMAIN 137 145 POLY-SER.
FT VARSPLIC 641 643 EST -> GHS (in isoform 2).
FT VARSPLIC 644 1327 /FTId=VSP_004538.
FT VARSPLIC 644 1327 Missing (in isoform 2).
FT MUTAGEN 1184 1184 H->A: LOSS OF ACTIVITY, WHEN ASSOCIATED
FT MUTAGEN 1291 1291 WITH A-1291.
FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY, WHEN ASSOCIATED
FT SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;
Query Match 11.8%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 5.2; Gaps 4;
Matches 35; Conservative 15; Mismatches 57; Indels 8;
QY 6 VAAFAAI-VVGSALAGVVPQMGVGGNGHNGSGDSTLSIYQVGSNAALALQSDAR 64
Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSSPSPSPGSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVGOGADNSITELTQNG--FNNNATIDQWNAKNSDI 113
Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
158 GVSSTAPLPGGAAGPGTGPVAVSGALRELLACRNGDVSERVKRLVDAAVNAKDM 212
RESULT 10
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR: B1041 OR Z1675 OR ECS14119.
OS Escherichia coli, and
OC Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RA "Expression of two csg operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
RT K-12";
RL Mol. Microbiol. 18:661-670(1995).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sanpei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map.";
[4]
RT RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RL MEDLINE=21074935; PubMed=11206551;
RN Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RC Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RX Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RN Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
[5]
RT RC STRAIN=O157:H7 / RIMD 0509952;
RL MEDLINE=21156231; PubMed=11258796;
RN Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RC Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RX Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
[6]
RT RC SEQUENCE OF 1-21 FROM N.A.
RL STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csgBA promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
of the nucleoid-associated protein H-NS";
RL Mol. Microbiol. 13:1021-1032(1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; X90754; CAA62281.1; -;
DR EMBL; AB000205; AAC74125.1; -;
DR EMBL; D90741; BAA35831.1; -;
DR EMBL; AB005315; AAG55787.1; -;
DR EMBL; AP002554; BAB34842.1; -;
DR PIR; C90806; C90806;
DR PIR; G85665; G85665;
DR PIR; S70787; S70787;
DR EcoGene; EG12621; csgB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 11.7%; Score 91.5; DB 1; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.48;
Matches 32; Conservative 17; Mismatches 54; Indels 7; Gaps 4;
QY 9 FBAIVVGSAL-AGVVPWGGGNGHNGSSGPDSTLSIY-QYGSANAALALQSDARK 65
DB 33 FAVNELSKSFNQAIH---GQAGTNNBAQLRQGGSKLLAVVQGGSSNRA-KIDQTGDY 89
QY 66 SETTITQCYGNGADVGGQADNSTIETQNGFRNATIDQWNAKNSDITV 115

DB 89 NLAYIDQAGSANDASISQAYGNTAMTIQKSGNKNITQYGTQKTAIVV 138
RESULT 11
VG38_BPT2
ID VG38_BPT2 STANDARD; PRT; 262 AA.
AC P07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283911; PubMed=3302276;
RA Riede I., Drexler K., Eschbach M.L., Henning U.;
RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of
bacteriophages T2, K3 and of K3 host range mutants";
RL J. Mol. Biol. 194:31-39(1987).
CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as
the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
AS RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR PIR; X05312; CAA28935.1; -;
DR PIR; S00275; S00275.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
KW Fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918P6C745 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 262;
Best Local Similarity 34.1%; Pred. No. 1.3;
Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;
QY 27 GGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSYGNGADVGGQAD 86
DB 175 GGGGRPFVGGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGNGVAGGG 225
QY 87 NSTIETQNGFRNATIDQWNAKNSDIT 114
DB 226 -----RWGKNVYTSGGAGAANTV 245
RESULT 12
MSA2_PLAF2
ID_MSA2_PLAF2 STANDARD; PRT; 347 AA.
AC Q03646;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface antigen 2 precursor (MSA-2).
GN MSA2.
OS Plasmodium falciparum (isolate Nig32 / Nigeria).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70150;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91156685; PubMed=2000383;
RA Smythe J.A., Coppel R.L., Day K.P., Martin R.K., Oduola A.M.J.,
RA Kemp D.J., Anders R.F.;


```

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davis K., Devlin J., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22266494; PubMed=1218036;
RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBooy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.J., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garner T., Eidmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -I- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY
CC CC
CC -I- CAUTION: Ref.l sequence differs from that shown due to a
CC frameshift in position 59. Ref.l sequence has been checked by
CC authors in Ref.1 and they report that no errors have been found.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR ENBL; Z73966; CA98228.1; ALT_FRAME.
DR ENBL; AE007065; AK46440.1; -.
DR ENBL; BX248341; CAD96978.1; -.
DR TIGR; MT2159; -.
DR TubercuList; RV2098c; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE_1.
DR ProDom; PD001223; PE_region_1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 312 312 G -> GG (IN REF. 1).
SQ SEQUENCE 491 AA; 41979 MW; 12CB630C59CA0C13 CRC64;
Query Match 11.3%; Score 88; DB 1; Length 491;
Best Local Similarity 26.8%; Pred. No. 3 6;
Matches 30; Conservative 13; Mismatches 39; Indels 30; Gaps 5;
QY 20 AGVPWVGCGGN-----HNGGNSGGPDSLTIIYQGSANAALALOSDARKSETTTQS 74
||| ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 222 AGLIHGGAGNCGDGCHGSGKAGSGSGGFQFGGAGLL----- 264
QY 75 YNGADVGGOG--NSTITQTGNFRNNATIDQNAKNSD---ITVGYQGN 121
||| ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 265 YNGNHAAGSGNGDAGTGVSDGP---AGLGSGRGGDAGLIIVGGCGGN 313

```

| | | | | |
|----|--|-----------|------|---------|
| ID | OMB2_NEIMS | STANDARD; | PRT; | 331 AA. |
| AC | P30688; | | | |
| DT | 01-APR-1993 (Rel. 25, Created) | | | |
| DT | 01-APR-1993 (Rel. 25, Last sequence update) | | | |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) | | | |
| DE | Major outer membrane protein P.IB precursor (Protein IB) (PIB) | | | |
| DE | (Porin) (Class 3 protein). | | | |
| GN | PORB. | | | |
| OS | Neisseria meningitidis (serogroup B). | | | |
| OC | Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; | | | |
| OC | Neisseriaceae; Neisseria. | | | |
| OX | NCBI_TaxID=491; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=CCUG 37604 / M981 / Serogroup B / Serotype 4; | | | |
| RX | MEDLINE=93051225; PubMed=1330818; | | | |
| RA | Ward M.J., Lambden P.R., Heckels J.E.; | | | |
| RT | "Sequence analysis and relationships between meningococcal class 3 | | | |
| RT | serotype proteins and other porins from pathogenic and non-pathogenic | | | |
| RT | Neisseria species."; | | | |
| RT | FEMS Microbiol. Lett. 73:283-289(1992). | | | |
| RL | [2] | | | |
| RN | SEQUENCE FROM N.A. | | | |
| RP | STRAIN=CU385 / Serogroup B / Serotype 4 / Subtype 15; | | | |
| RC | MEDLINE=93116587; PubMed=1335540; | | | |
| RX | Zapata G.A., Vann W.F., Rubinstein Y., Frasch C.E.; | | | |
| RA | "Identification of variable region differences in Neisseria | | | |
| RT | meningitidis class 3 protein sequences among five group B | | | |
| RT | serotypes."; | | | |
| RT | Mol. Microbiol. 6:3493-3499(1992). | | | |
| RL | [1] | | | |
| CC | -!- FUNCTION: Serves as a slightly cation selective porin. | | | |
| CC | -!- SUBUNIT: Homotrimer. | | | |
| CC | -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane. | | | |
| CC | -!- SIMILARITY: Belongs to the Gram-negative porin family. | | | |
| CC | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | | | | |
| CC | EMBL; X65531; CAA46501.1; - | | | |
| DR | PIR; S21409; S21409. | | | |
| DR | PIR; S28441; S28441. | | | |
| DR | InterPro; IPR001702; Porin Gram-ve. | | | |
| DR | Pfam; PF00267; Gram-ve porins; 1. | | | |
| DR | PRINTS; PR00182; ECOLNEIPORIN. | | | |
| DR | PROSITE; PS00576; GRAM_NEG_PORIN; 1. | | | |
| KW | Outer membrane; Porin; Transmembrane; Antigen; Signal. | | | |
| FT | SIGNAL 1 19 | | | |
| FT | CHAIN 20 331 | | | |
| FT | VAR 70 71 | | | |
| FT | NG -> KR (IN STRAIN CU385). | | | |
| FT | SEQUENCE 331 AA; 35741 MW; 35EA35B7EBD28301 CRC64; | | | |
| QY | Query Match 11.2%; Score 87; DB 1; Length 331; | | | |
| Db | Best Local Similarity 25.8%; Pred. No. 2.8; 78; Indels 22; Gaps 6; | | | |
| Db | Matches 42; Conservative 23; Mismatches 22; Gaps 6; | | | |
| QY | 4 LKVAFAAIVVSGSALAGVVPWGNGGNGHNGGNSGPDSTLSIYVYGS-----A 53 | | | |
| Db | 13 LPVAAMADVTLTYIKAGV--ETSRVHGNGQVSVETGT-GIVDLGSKIGFKGQEDIG 69 | | | |
| QY | 54 NAALALQSDARKSETTITQSGYCN-----GADVCGQADNSTIETLTCNGFRNNATIDKNA 108 | | | |
| Db | 70 NGLKAIWQVEQKASIACTGDSGNGNQSFGLKGGFG--KLVRGLNSVLKDTGIDNPMD 127 | | | |
| QY | 109 KNSDITVGYGGGNAAL--VNYQLVTRVVVTHMAHANATANQY 151 | | | |
| Db | 128 KSDYLGYNKTAEPFARLISVRYDSPFAGLSGVSQVYALNDNAGKY 172 | | | |

| | | | | |
|-----------|--|------|----------|--|
| RESULT 15 | | | | |
| PER_DROWI | STANDARD; | PRT; | 1093 AA. | |
| ID | Q03297; O18421; O18422; P91721; P91722; | | | |
| AC | Q03297; O18421; O18422; P91721; P91722; | | | |
| DT | 01-APR-1993 (Rel. 25, Created) | | | |
| DT | 01-JUL-1998 (Rel. 36, Last sequence update) | | | |
| DT | 15-OCT-2001 (Rel. 40, Last annotation update) | | | |
| DE | Period circadian protein (Fragment). | | | |
| GN | PER. | | | |
| OS | Drosophila willistoni (Fruit fly). | | | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | | |
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | | |
| OX | NCBI_TaxID=7260; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=Various strains; | | | |
| RX | MEDLINE=97357421; PubMed=9214747; | | | |
| RA | Gleason J.M., Powell J.R.; | | | |
| RT | "Interspecific and intraspecific comparisons of the period locus in | | | |
| RT | the Drosophila willistoni sibling species."; | | | |
| RL | Mol. Biol. Evol. 14:741-753(1997). | | | |
| RN | [2] | | | |
| RP | SEQUENCE OF 579-646 FROM N.A. | | | |
| RC | MEDLINE=93196482; PubMed=8450754; | | | |
| RX | Peixoto A.A., Campesan S., Costa R.H., Kyriacou C.P.; | | | |
| RA | "Molecular evolution of a repetitive region within the per gene of | | | |
| RT | Drosophila."; | | | |
| RL | Mol. Biol. Evol. 10:127-139(1993). | | | |
| CC | -!- FUNCTION: Essential for biological clock functions. Determines the | | | |
| CC | period length of circadian and ultradian rhythms; an increase in | | | |
| CC | PER dosage leads to shortened circadian rhythms and a decrease in | | | |
| CC | leads to lengthened circadian rhythms. Essential for the circadian | | | |
| CC | rhythmicity of locomotor activity, eclosion behavior, and for the | | | |
| CC | thoracic nervous system. The biological cycle depends on the | | | |
| CC | rhythmic formation and nuclear localization of the TIM-PER | | | |
| CC | complex. Light induces the degradation of TIM, which promotes | | | |
| CC | elimination of PER. Nuclear activity of the heterodimer | | | |
| CC | coordinatively regulates PER and TIM transcription through a | | | |
| CC | negative feedback loop. Behaves as a negative element in circadian | | | |
| CC | transcriptional loop. Does not appear to bind DNA, suggesting | | | |
| CC | indirect transcriptional inhibition (By similarity). | | | |
| CC | -!- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then | | | |
| CC | translocates into the nucleus (By similarity). | | | |
| CC | -!- SUBCELLULAR LOCATION: Nuclear at specific periods of the day. | | | |
| CC | First accumulates in the perinuclear region about one hour before | | | |
| CC | translocation into the nucleus. Interaction with TIM is required | | | |
| CC | for nuclear localization (By similarity). | | | |
| CC | -!- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE | | | |
| CC | DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN | | | |
| CC | THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER | | | |
| CC | PER-TIM (BY SIMILARITY). | | | |
| CC | -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATULXCO. | | | |
| CC | -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains. | | | |
| CC | -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain. | | | |
| CC | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | | | | |
| CC | EMBL; U51055; AAB41360.1; - | | | |
| DR | EMBL; U51056; AAB41361.1; - | | | |
| DR | EMBL; U51057; AAB41362.1; - | | | |
| DR | EMBL; U51058; AAB41363.1; - | | | |
| DR | EMBL; U51059; AAB41364.1; - | | | |
| DR | EMBL; U51060; AAB41365.1; - | | | |
| DR | EMBL; U51061; AAB41366.1; - | | | |
| DR | EMBL; U51062; AAB41367.1; - | | | |

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVVTHEMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP_TREMBL_25:*
- 2: sp_archaea:*
- 3: sp_bacteria:*
- 4: sp_fungi:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mbc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 686 | 88.1 | 152 | 2 | O33802 salmonella |
| 2 | 591.5 | 75.9 | 150 | 2 | Q7X243 |
| 3 | 537 | 68.9 | 149 | 2 | Q7X240 citrobacter |
| 4 | 495.5 | 63.6 | 152 | 16 | Q8CW63 |
| 5 | 427.5 | 54.9 | 150 | 2 | Q7X237 enterobacte |
| 6 | 385 | 49.4 | 76 | 2 | Q54069 salmonella |
| 7 | 122 | 15.7 | 29 | 2 | Q9S3J5 |
| 8 | 113.5 | 14.6 | 3501 | 16 | Q8Y106 |
| 9 | 113.5 | 14.6 | 3552 | 16 | Q8XSD6 |
| 10 | 108.5 | 13.9 | 191 | 10 | Q7XDR3 |
| 11 | 107 | 13.7 | 502 | 16 | Q8EIH4 |
| 12 | 106 | 13.6 | 1748 | 5 | Q4821 |
| 13 | 105 | 13.5 | 2035 | 2 | Q9XCJ4 |
| 14 | 105 | 13.5 | 2039 | 16 | Q8ZNS7 |
| 15 | 104.5 | 13.4 | 1286 | 2 | Q841Y5 campylobact |
| 16 | 104 | 13.4 | 2174 | 16 | Q92U08 rhizobium m |

| | | | | | |
|----|-------|------|------|----|--------------------|
| 17 | 103 | 13.2 | 139 | 16 | Q8EIH3 |
| 18 | 102.5 | 13.2 | 624 | 3 | Q8NIV1 |
| 19 | 102 | 13.1 | 1422 | 16 | Q8EFU3 |
| 20 | 100.5 | 12.9 | 152 | 2 | Q7X241 |
| 21 | 100 | 12.8 | 179 | 2 | Q33801 salmonella |
| 22 | 99.5 | 12.8 | 151 | 2 | Q7X244 citrobacter |
| 23 | 99 | 12.7 | 1765 | 16 | Q7V8S5 |
| 24 | 98.5 | 12.6 | 1222 | 5 | Q9W4F0 |
| 25 | 98.5 | 12.6 | 1222 | 5 | Q8T8L9 |
| 26 | 98.5 | 12.6 | 1351 | 5 | Q8SX56 |
| 27 | 98.5 | 12.6 | 1354 | 5 | Q8MPN4 |
| 28 | 98.5 | 12.6 | 7716 | 16 | Q7UWZ8 |
| 29 | 97.5 | 12.5 | 154 | 16 | Q89J15 |
| 30 | 97.5 | 12.5 | 348 | 13 | Q93397 |
| 31 | 97.5 | 12.5 | 739 | 2 | Q9X687 |
| 32 | 97.5 | 12.5 | 3659 | 16 | Q98LN6 |
| 33 | 97 | 12.5 | 1618 | 2 | Q9KXB1 |
| 34 | 96.5 | 12.4 | 151 | 2 | Q7X238 |
| 35 | 96 | 12.3 | 145 | 16 | Q8U6N9 |
| 36 | 96 | 12.3 | 157 | 16 | Q88HG0 |
| 37 | 95.5 | 12.3 | 130 | 16 | Q89J14 |
| 38 | 95.5 | 12.3 | 453 | 5 | Q9N6M8 |
| 39 | 95.5 | 12.3 | 1615 | 2 | Q9KXA8 |
| 40 | 95 | 12.2 | 362 | 16 | Q8EV84 |
| 41 | 95 | 12.2 | 1613 | 2 | Q84OU5 |
| 42 | 95 | 12.2 | 1618 | 2 | Q9KXB4 |
| 43 | 94.5 | 12.1 | 160 | 16 | Q8CW64 |
| 44 | 94.5 | 12.1 | 453 | 5 | Q9NGF6 |
| 45 | 94.5 | 12.1 | 453 | 5 | Q9NGF7 |

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.

AC O33802;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AcfA protein (fragment).
GN AcfA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells."
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL AJ000514; CAA04151.1; -.
FT NON_TER
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 88.1%; Score 686; DB 2; Length 152;

Best Local Similarity 88.7%; Pred. No. 4.1e-45;

Matches 134; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGNGNGSGPDSSTLSIVYGSANAALAIQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGNGNGSGPDSSTLSIVYGSANAALAIQ 60

QY 61 SPARKSETTITGSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVQYGG 120

Db 61 SPARKSETTITGSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151

```

Db      121 NNAALVNTQASDSSVVMVQVGFNNAPANQY 151
      ||||| : : : |||||
      ||||| PRELIMINARY; PRT; 150 AA.

RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 75.9%; Score 591.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 6.7e-38;
Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQW-GGGNGHGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLQVAAPAAIIVVSGSALAGVVPQWGGGG-GGGSSSGSPSTLSIYQGSVNNAAALQ 59
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGFRRNATIDQWNAKNSDITVGYGG 120
DB 60 SDARKSETTITQHGFGNGADVGQSDNSTIDITQGFKNATIDQWNGKNSDITVSYGG 119

QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
DB 120 HNAALVNTQASDSSVLVHQVGFNNATANQY 150
      ||||| : : : |||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 68.9%; Score 537; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 9.7e-34;
Matches 108; Conservative 17; Mismatches 24; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGNGHGGNSGPDSTLSIYQGSANAALALQ 60
      ||||| : : : |||||

```

```

Db      1 MKLLKVAAPAAIIVVSGSALAGVVPQW--GGNHGGNSGPDSSLSIYQGSNNSANAALQ 58
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGFRRNATIDQWNAKNSDITVGYGG 120
      ||||| : : : |||||
      ||||| DB 59 SDARKSDVTITQHGRCNGAVVGGGADDTISLKTQTFQNSATIDQWNAKNAIDISVTQFGG 118
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
      ||||| : : : |||||
      ||||| DB 119 NGALVNTQASDSSNVLIQVGFNNATANQH 149

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 63.6%; Score 495.5; DB 16; Length 152;
Best Local Similarity 65.1%; Pred. No. 1.5e-30;
Matches 99; Conservative 21; Mismatches 31; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQW-GGGNGHGGNSGPDSTLSIYQGSANAALAL 59
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQYGGGGNHGGNGNSGPNSELNIYQGGNSALAQ 60
      ||||| : : : |||||
      ||||| QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQGFRRNATIDQWNAKNSDITVGYGG 119
      ||||| DB 61 QADARNSDLTITQHGNGADVGQSDSDSIDITQGFNGSATIDQWNGKDSMTVYKQFG 120
QY 120 GNNALVNYDQLVTRVVTHEMAHANNATANQY 151
      ||||| : : : |||||
      ||||| DB 121 GGNGAIVDQTASNSVNVTVQVGFNNATANQY 152

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 508BB2D872DF15F3 CRC64;

Query Match 54.9%; Score 427.5; DB 2; Length 150;
Best Local Similarity 58.3%; Pred. No. 2.3e-25;
Matches 88; Conservative 26; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVPGWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKFIKVAALAAIVVSGSAGMAGNIQ-GGWGHGHGGYGGPNSTLNIYQNGGNSALQ 59

QY 61 SDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 60 TDARNSVLNISOTGGGADYQGGSDSSINLTQNGFNSATLQWNSKDSVMNVSYGG 119

QY 121 NNAALVNDQVTRVTHEMAHANNATANOY 151
DB 120 LNALVDQTSASNTVNTQIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae."
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCE1 CRC64;

Query Match 49.4%; Score 385; DB 2; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.9e-22;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GNHNGGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADYQGGADNST 89
DB 1 GNHNGGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADYQGGADNST 60

QY 90 IELTQNGFRNNATIDQ 105
DB 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q53J5 PRELIMINARY; PRT; 29 AA.
AC Q53J5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAY-2001 (TRENBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:H80 isolates associated with
RT ISI insertions in csgh and reduced persistence in poultry infection."
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -.
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0078; 2; Indels 0; Gaps 0;
Matches 26; Conservative 1; Mismatches 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29

RESULT 8
Q8Y106 PRELIMINARY; PRT; 3501 AA.
ID Q8Y106;
AC Q8Y106;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RSC06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21691879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cumnac S., Denange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schliex F.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AF646061; CAD14589.1; -.
DR GO: GO:0004519; F:endonuclease activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001604; Endonuclease.
DR InterPro: IPR008619; Fil haemagg.
DR InterPro: IPR008638; Haemagg act.
DR Pfam: PF05594; Fil haemagg act.
DR Pfam: PF05860; Haemagg act; 1.
DR PROSITE: PS01070; NUCLEASE_NON_SPEC; 1.
SQ Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.6%; Score 113.5; DB 16; Length 3501;
Best Local Similarity 28.8%; Pred. No. 8.3;
Matches 36; Conservative 19; Mismatches 43; Indels 27; Gaps 4;

QY 15 SGSALAGVVPQWGGGNGHNGG-NSSGPDSTLSIYQYGSANAA----- 56
DB 2431 SGSHFTAGPSWGLGRNVGGGPNSSG---VGLAFYGAHSAADNAAGNSRQNASVWIG 2486

QY 57 LALQSDARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVIG 116
DB 2487 KSVQVQARTGDIIVSGSGISALSDVLLAKQKQKVDIVAGNDTSSRHD-----HSDRTIG 2541

QY 117 QYGN 121
|||

Page 4

Qy 122 N 122
Db 350 N 350

RESULT 12
Q94821 PRELIMINARY; PRT; 1748 AA.
ID Q94821 P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
AC Q94821 P92146; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CNUB protein.
GN CNUB

OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymena; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=3357771;
RA Martindale D.W., Taylor F.M.;
RT "Multiple introns in a conjugation-specific gene from Tetrahymena
thermophila";
RL Nucleic Acids Res. 16:2189-2201(1988).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=94051569; PubMed=8233798;
RA Taylor F.M., Martindale D.W.;
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
encoded by cnjB, a Tetrahymena gene active during meiosis.";
RL Nucleic Acids Res. 21:4610-4614(1993).
DR EMBL; X06462; CAB37323.1; -;
DR EMBL; L03710; AAC37171.1; -;
DR PIR; S42136; S42136.
DR HSSP; P05888; 1A4F.
GO; GO:0003676; P:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf CCHC.
DR Pfam; PF00098; zf-CCHC; 7.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SMO0343; Znf_C2HC; 7.
DR PROSITE; PS0158; ZF_CCHC; 7.
FT CONFLICT 251 251 M -> I (IN REF. 1).
FT CONFLICT 256 256 I -> N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 13.6%; Score 106; DB 5; Length 1748;
Best Local Similarity 32.1%; Pred. No. 14;
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;

Qy 25 QWGGGNGHGG---GNSSGPDSTLSIYQYGSANAALALQSPARKSETTIT---QSGYGN 77
Db 1640 QFGGGNGNGGSGWTSSGSDN-----CQSNVQSTTTSSGGWSSGSGN 1685

Qy 78 GADYQGGADNSTIETLQNGFRNNATIDOWNARNSDITVQY--YGNNA 124
Db 1686 QTGGGWSGNDN-----QQQNGENTGGGWSGNSQNTNNESSGWSNQA 1729

RESULT 13
Q9XCJ4 PRELIMINARY; PRT; 2035 AA.
ID Q9XCJ4
AC Q9XCJ4
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ShdA.
GN SHDA.

OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC14028;
RA Kingsley R.A., van Amsterdam K., Baumber A.J.;
RT "The presence of a pathogenicity island specific to Salmonella
enterica subspecies I correlates with adaptation to warm blooded
animals"; (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC14028;
RA Kingsley R.A., van Amsterdam K., Edwards E.W., Hargis B.M.,
Baumber A.J.;
RT "Complete sequence of the xseA-his intergenic region of the S.
enterica serotype Typhimurium genome and its distribution within the
genus Salmonella";
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; A1140550; AAD25110.2; -;
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR004899; Pertactin.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF03212; Pertactin; 1.
DR TIGRFAMS; TIGR01414; autotrans_bar1; 3.
DR PROSITE; PSC0584; PFKB_KINASES_2; 2.
SQ SEQUENCE 2035 AA; 207032 MW; 295DB82FFCA84FAB CRC64;

Query Match 13.5%; Score 105; DB 2; Length 2035;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;

Qy 10 AAIYVSGSALAG-----VVPQGGGNGHGGG---NSSGPD----- 42
Db 90 AAIYVSGVATVWQPTTGTGLVETSGGADDPGGKYVSNALSDHVAILELTDAKI 149

Qy 43 STLSIYQYGSANA-----ALALQSPARKS-ETTITQSGYNGADVG 82
Db 150 TTTGIYQGISAADGSTLRLDSTLIDGNFGVMTLYTGEATLDGTIVEAANSSAQVQ 209

Qy 83 QGA-----DNSTIELTQ---NGFRNNATIDOWNARNSDITVQYGVG-----GNNAAALVN 127
Db 210 QGSTNLVLDGSTITLAQQGVVAGTATDEGSTLNLSDSVSSAGTMTIOGTNKALN 269

Qy 128 YDOLVTRVVTHEMA-----HANNAT 147
Db 270 ---LTNATITHNAGAAVQANNAT 291

RESULT 14
Q8ZN57 PRELIMINARY; PRT; 2039 AA.
ID Q8ZN57
AC Q8ZN57;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to the C-terminal region of AIDA, IcsA, subspecies I specific,
Peyer's patch colonization and shedding factor.
GN SHDA OR STM2513.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Van E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
Lt2.";

```

RL Nature 413:852-856 (2001).
DR EMBL: AEO08813; AAL21407.1; -.
DR InterPro: IPR006315; Autotransport.
DR InterPro: IPR005546; Autotransporter.
DR InterPro: IPR004899; Pertactin.
DR InterPro: IPR002173; PfkB.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRFAMs: TIGR01414; autotrans_barl; 3.
DR PROSITE: PS00584; PFKB_KINASES_2; 2.
KW Complete Proteome.
SQ SEQUENCE 2039 AA; 207127 MW; 894E41F8F29339EA CRC64;

Query Match 13.5%; Score 105; DB 16; Length 2039;
Best Local Similarity 26.3%; Pred. No. 20;
Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;

QY 10 AAIIVSGSALAG-----VVPQGGGGHNGGG-----NSSGPD----- 42
DB 94 AALVYSGVATVGQPTTIVTGTGLVETSGGADDPGGKYVSNALSLDHYAILELTDARI 153
QY 43 STLISIVQGSANA-----ALALQSDARKS-ETTITQSGYNGADV 82
DB 154 TTGIVTQGISADGSLTBLTSTLTIDNFGVWTLTGSEATLDGTIVEANSSAQVQ 213

QY 83 QGA-----DNSTIELTQ---NGFRNNATIDQNAKN-SDITVQYQV-----GNNAALVN 127
DB 214 QGSTLVLDGSTITLAQQGQINVVAGNTATDEGSTLNLSDSVSSAGTMTSTIQTGINKAALN 273

QY 128 YQQLVTRVVTHEMA-----HANNAT 147
DB 274 ----LTNATITHTNAGAAVQANAT 295

RESULT 15
Q841Y5 PRELIMINARY; PRT; 1286 AA.
ID Q841Y5;
AC Q841Y5;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DE Putative high-molecular-weight surface-exposed protein Cf0009.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=90354448; PubMed=2387868;
RA Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure.";
RL J. Biol. Chem. 265:14529-14535 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=9103477; PubMed=2229082;
RA Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure.";
RL J. Biol. Chem. 265:19372-19372 (1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=92394895; PubMed=1522068;
RA Tummuru M.K., Blaser M.J.;
RT "Characterization of the Campylobacter fetus sapA promoter: evidence
that the sapA promoter is deleted in spontaneous mutant strains.";
RL J. Bacteriol. 174:5916-5922 (1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;

```

```

RX MEDLINE=93348254; PubMed=8346244;
RA Tummuru M.K., Blaser M.J.;
RT "Rearrangement of sapA homologs with conserved and variable regions in
Campylobacter fetus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7265-7269 (1993).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=95204338; PubMed=7896695;
RA Dworkin J., Tummuru M.K., Blaser M.J.;
RT "A lipopolysaccharide-binding domain of the Campylobacter fetus S-
layer protein resides within the conserved N terminus of a family of
silent and divergent homologs.";
RL J. Bacteriol. 177:1734-1741 (1995).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=95059317; PubMed=9851986;
RA Thompson S.A., Shedd O.L., Ray K.C., Beins M.H., Jorgensen J.P.,
Blaser M.J.;
RT "Campylobacter fetus surface layer proteins are transported by a type
I secretion system.";
RL J. Bacteriol. 180:6450-6458 (1998).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX PubMed=12694614;
RA Tu Z.C., Wassenaar T.M., Thompson S.A., Blaser M.J.;
RT "Structure and genotypic plasticity of the Campylobacter fetus sap
locus.";
RL Mol. Microbiol. 48:685-698 (2003).
DR EMBL: AY211269; AAC064216.1; -.
DR InterPro: IPR005546; Autotransporter.
DR Pfam: PF03797; Autotransporter; 1.
SQ SEQUENCE 1286 AA; 134079 MW; A1FF9CBG34158789 CRC64;

Query Match 13.4%; Score 104.5; DB 2; Length 1286;
Best Local Similarity 27.8%; Pred. No. 13;
Matches 44; Conservative 26; Mismatches 57; Indels 31; Gaps 9;

QY 5 KVAFAAIVVSGSALAGVVPQGGGGHNGGNS---SGPDTLSIYQVGSANAALALQS 61
DB 376 QVASENLVVISGGTIN--VPTIGGGSATNATNNQVTISGKVTSTSYGNNAN-----K 427
QY 62 DARKSETTITQSGYNGADV--GQAGDNSTIELTQNGFNATIDQNAKNSDITVQYQ 119
DB 428 SANENKVITE-GTANVADIYGKSISSNSI-----ANKNSITISGGLQVTNI-----YG 477
QY 120 GNNAALVNYDQL-----VTRVV-THEMAHANNATAN 149
DB 478 GHSKADANSIQISNGGNNINIVGGHAQDHTNLNTIN 515

Search completed: August 2, 2004, 14:54:36
Job time : 31.7 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVWVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1596107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 775 | 100.0 | 151 | 3 AAB36347 | Aab36347 Agfa::PT3 |
| 2 | 714 | 92.1 | 151 | 3 AAB36352 | Aab36352 Agfa::PT3 |
| 3 | 696 | 89.8 | 151 | 3 AAB36346 | Aab36346 Agfa::PT3 |
| 4 | 693 | 89.4 | 151 | 2 AAR74625 | Aar74625 Agfa sequ |
| 5 | 693 | 89.4 | 151 | 3 AAB36341 | Aab36341 Salmonell |
| 6 | 692 | 89.3 | 151 | 2 AAW23570 | Aaw23570 Salmonell |
| 7 | 659 | 85.0 | 151 | 3 AAB36355 | Aab36355 Agfa::PT3 |
| 8 | 612 | 79.0 | 151 | 3 AAB36353 | Aab36353 Agfa::PT3 |
| 9 | 611 | 78.8 | 151 | 3 AAB36349 | Aab36349 Agfa::PT3 |
| 10 | 609 | 78.6 | 151 | 3 AAB36350 | Aab36350 Agfa::PT3 |
| 11 | 603 | 77.8 | 151 | 3 AAB36354 | Aab36354 Agfa::PT3 |
| 12 | 602 | 77.7 | 151 | 3 AAB36351 | Aab36351 Agfa::PT3 |
| 13 | 578 | 74.6 | 151 | 3 AAB36348 | Aab36348 Agfa::PT3 |
| 14 | 528 | 68.1 | 151 | 3 AAB36343 | Aab36343 Escherich |
| 15 | 523 | 67.5 | 151 | 7 ABR82651 | AbR82651 E. coli C |
| 16 | 514 | 66.3 | 120 | 2 AAR62761 | Aar62761 Agfa sequ |
| 17 | 514 | 66.3 | 120 | 2 AAW23369 | Aaw23369 Salmonell |
| 18 | 463 | 59.7 | 142 | 2 AAR52664 | Aar52664 Fibronect |
| 19 | 391 | 50.6 | 122 | 2 AAR52663 | Aar52663 FNB curli |
| 20 | 237 | 30.6 | 45 | 3 AAB36316 | Aab36316 Salmonell |
| 21 | 132 | 17.0 | 22 | 3 AAB36318 | Aab36318 Salmonell |
| 22 | 123 | 15.9 | 23 | 3 AAB36321 | Aab36321 Salmonell |
| 23 | 123 | 15.9 | 23 | 3 AAB36326 | Aab36326 Salmonell |
| 24 | 123 | 15.9 | 23 | 3 AAB36338 | Aab36338 Salmonell |
| 25 | 115 | 14.8 | 22 | 3 AAB36325 | Aab36325 Salmonell |

| | | | | | |
|----|-------|------|------|------------|--------------------|
| 26 | 115 | 14.8 | 22 | 3 AAB36339 | Aab36339 Salmonell |
| 27 | 115 | 14.8 | 22 | 3 AAB36320 | Aab36320 Salmonell |
| 28 | 113 | 14.6 | 24 | 7 ABR82644 | AbR82644 E. coli C |
| 29 | 109 | 14.1 | 23 | 3 AAB36340 | Aab36340 Salmonell |
| 30 | 109 | 14.1 | 23 | 3 AAB36324 | Aab36324 Salmonell |
| 31 | 109 | 14.1 | 23 | 3 AAB36319 | Aab36319 Salmonell |
| 32 | 104.5 | 13.5 | 151 | 3 AAB36344 | Aab36344 Escherich |
| 33 | 98.5 | 12.7 | 151 | 3 AAB36342 | Aab36342 Salmonell |
| 34 | 98 | 12.6 | 26 | 7 ABR82645 | AbR82645 E. coli C |
| 35 | 95 | 12.3 | 186 | 6 ABU21488 | AbU21488 Protein e |
| 36 | 93 | 12.0 | 19 | 3 AAB36323 | Aab36323 Salmonell |
| 37 | 93 | 12.0 | 19 | 3 AAB36326 | Aab36326 Salmonell |
| 38 | 93 | 12.0 | 19 | 3 AAB36328 | Aab36328 Salmonell |
| 39 | 92 | 11.9 | 23 | 3 AAB36331 | Aab36331 Escherich |
| 40 | 92 | 11.9 | 673 | 3 AAY44403 | Aay44403 Human tru |
| 41 | 92 | 11.9 | 673 | 5 AAU79538 | Aau79538 Truncated |
| 42 | 92 | 11.9 | 949 | 3 AAY44404 | Aay44404 Human tru |
| 43 | 92 | 11.9 | 949 | 5 AAU79539 | Aau79539 Truncated |
| 44 | 92 | 11.9 | 1327 | 3 AAY44402 | Aay44402 Human tan |
| 45 | 92 | 11.9 | 1327 | 3 AAB27212 | Aab27212 Human tan |

ALIGNMENTS

RESULT 1
AAB36347
ID AAB36347 standard; protein; 151 AA.
XX
AC AAB36347;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX N-PSDB; AAC64623.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 5.4e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 QY 121 LVTRVVTHEMAHSAVMVQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHSAVMVQVGFNNATANQY 151

RESULT 2
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 XX
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PR protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (J) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC -Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX

SQ Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 4.3e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
 QY 121 LVTRVVTHEMAHSAVMVQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHSAVMVQVGFNNATANQY 151

RESULT 3
 AAB36346
 ID AAB36346 standard; protein; 151 AA.
 XX
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64622.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 89.8%; Score 696; DB 3; Length 151;
 Best Local Similarity 87.6%; Pred. No. 2.4e-59;
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQVGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQVGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETLQNGFRNNATIDQWNAKNSDITVGYG 117
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETLQNGFRNNATIDQWNAKNSDITVGYG 120
 QY 118 -----YDILVTRVVTTHMAHASVMVROVGFNNATANQY 151
 DB 121 NNAALVNYDQVTRVVTTHMAHA-----NNATANQY 151
 RESULT 4
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX Agfa sequence.
 DE
 XX Salmonella; Agfa; vaccine.
 KW
 XX Salmonella.
 OS

PN WO9425598-A2.
 XX
 XX 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 XX
 XX 26-APR-1993; 93US-00054452.
 XX
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
 PI
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ37467.
 DR
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 XX
 XX Disclosure; Fig 7B; 95pp; English.
 PS
 XX The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 CC
 XX Sequence 151 AA;
 SQ
 Query Match 89.4%; Score 693; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.7e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQVGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQVGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETLQNGFRNNATIDQWNAKNSDITVGYG 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETLQNGFRNNATIDQWNAKNSDITVGYG 120
 QY 121 LVTRVVTTHMAHASVMVROVGFNNATANQY 151
 DB 121 NNAALVNYDQVTRVVTTHMAHA-----NNATANQY 151
 RESULT 5
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 XX
 XX AAB36341;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 DE
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS
 XX WO2000060102-A2.
 PN
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64617.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CagA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 89.4%; Score 693; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.7e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
 DB 121 NNPAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 6
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 AC AAW23570;
 XX
 XX 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE *Salmonella enteritidis* 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.
 OS *Salmonella enteritidis*.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 123
 FT /note= "Encoded by GCC"
 XX

PN US5635617-A.
 XX 03-JUN-1997.
 XX 26-APR-1994; 94US-00233788.
 XX 26-APR-1993; 93US-00054452.
 XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX Collinson SK, Kay WW, Doran JL;
 XX WPI; 1997-309886/28.
 XX N-PSDB; AAT74142.
 XX Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 XX Example 2; Fig 7; 85pp; English.
 CC The present sequence represents agfa encoded by the full agfa gene
 CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of *Salmonella*
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 151 AA;
 Query Match 89.3%; Score 692; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 5.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
 DB 121 NNPAALVNQATSDSSVMVRQVGFNNATANQY 151
 RESULT 7
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 AC AAB36355;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 XX
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 XX Synthetic.
 XX
 XX WO2000060102-A2.
 PN
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX

```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64631.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 139; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 85.0%; Score 659; DB 3; Length 151;
XX Best Local Similarity 81.9%; Pred. No. 9.1e-56;
XX Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
XX
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHA-----SVMVROVGFGNNTANQY 151
DB 106 LVTRVVTHEMAHAGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
XX
RESULT 8
ID AAB36353 standard; protein; 151 AA.
XX AAB36353;
XX AC
XX XX
XX 26-FEB-2001 (first entry)
XX DE AgfA: P33#8 amino acid sequence SEQ ID NO:26.
XX XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX KW vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PF 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 79.0%; Score 612; DB 3; Length 151;
XX Best Local Similarity 81.5%; Pred. No. 3.2e-51;
XX Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
XX
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
DB 61 LVTRVVTHEMAHAGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVMVROVGFGNNTANQY 151
DB 121 NNAALVNQTASDSSVMVROVGFGNNTANQY 151
XX
RESULT 9

```

AAB36349
 XX AAB36349 standard; protein; 151 AA.
 AC AAB36349;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT#4 amino acid sequence SEQ ID NO:18.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64625.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.8%; Score 611; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4e-51;
 Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGGNGSSGPDSTLSITVQVGSAAVALAQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGGNGSSGPDSTLSITVQVGSAAVALAQ 60

QY 61 SDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 LVTRVVTHEMAHASVMVRQVFGNNTANQY 151
 DB 121 NNAALVNTQATSDSSVMVRQVFGNNTANQY 151

RESULT 10
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA

White AP, Doran JL, Collison SK, Kay WW;
 WPI; 2000-672631/65.
 N-PSDB; AAC64626.
 Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 which encodes foreign epitope or antigen, expresses recombinant Agfa
 protein useful for eliciting immune response in animal.
 Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 assembly system of strains of Salmonella, Escherichia coli and
 Enterobacteriaceae for the production of fimbriae comprising recombinant
 Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 directing recombination of a recombinant gene into the chromosome of the
 homologous species; (3) directing recombination of a recombinant gene
 back into the chromosome of the homologous species, replacing the native
 copy of that gene; and (4) eliciting an immune response in an animal,
 comprising separating an amino acid polymer comprising a recombinant Agfa
 protein containing a replacement segment or segments of foreign amino
 acid sequence or sequences grown on a Salmonella, E. coli or
 Enterobacteriaceae host cell, from the host cell and introducing the
 polymer into the animal in conjunction with a carrier or diluent. (I) is
 useful for the expression of recombinant Agfa protein which is useful for
 eliciting an immune response in an animal. In a fimbrial presentation
 system the heterologous antigens are presented in high numbers (up to
 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 immunogenicity and adhesion properties relevant for an efficient live
 vaccine, the carrier fimbrial subunit proteins are usually strong
 immunogens, which may be important for directing an immune response
 against the inserted epitope, and hybrid fimbriae are easy and
 inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 78.6%; Score 609; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 6.3e-51;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKYDQVTRVVTHEMAHSAVWVRQVGFNNATANQY 151

QY 121 LVTRVVTHEMAHSAVWVRQVGFNNATANQY 151
DB 121 NNAALVNGTASDSSVWVRQVGFNNATANQY 151

RESULT 11
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
AC AAB36354;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64630.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 77.8%; Score 603; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.4e-50;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SDARKSETTITQSGYNGGADYDQVTRVVTHEMAHSAVWVRQVGFNNATANQY 151

QY 121 LVTRVVTHEMAHSAVWVRQVGFNNATANQY 151
DB 121 NNAALVNGTASDSSVWVRQVGFNNATANQY 151

RESULT 12
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64627.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;

Query Match 77.7%; Score 602; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120

QY 121 LVTRVVTHEMAHASVVMVROVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 13
 ID AAB36348
 XX AAB36348 standard; protein; 151 AA.
 AC AAB36348;
 XX
 XX 26-FEB-2001 (first entry)
 DE
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNTV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW,
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;

Query Match 74.8%; Score 578; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 6.3e-48;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVVMVROVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVROVGFNNATANQY 151

RESULT 14
 ID AAB36343
 XX AAB36343 standard; protein; 151 AA.
 AC AAB36343;
 XX
 XX 26-FEB-2001 (first entry)
 DE
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Escherichia coli.
 OS WO2000060102-A2.
 XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sf17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 68.1%; Score 528; DB 3; Length 151;
Best Local Similarity 68.9%; Pred. No. 4.3e-43;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNGSGPDSLTISIYQGSANAALALQ 60
Db 1 MKLLKVAATAIVFSGSALAGVVPQYGGGNGHGGNGSGNSLNIIYQGGNSALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
Db 61 TDARNSDLTITQGGGNGADVGCGSDSSIDLITQRFNGSATLDQNGKNSMTVQFGG 120

Qy 121 LVTRVVVTHEMAHASVVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
ABR82651
ID ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
XX Escherichia coli.
XX
PN WO2003064446-A2.
XX
PD 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-BP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
PI
XX WPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
SQ Sequence 151 AA;

Query Match 67.5%; Score 523; DB 7; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.3e-42;
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNGSGPDSLTISIYQGSANAALALQ 60
Db 1 MKLLKVEATAIVFSGSALAGVVPQYGGGNGHGGNGSGNSLNIIYQGGNSALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
Db 61 TDARNSDLTITQGGGNGADVGCGSDSSIDLITQRFNGSATLDQNGKNSMTVQFGG 120

Qy 121 LVTRVVVTHEMAHASVVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

Search completed: August 2, 2004, 14:48:24
Job time : 44.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-14
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASWVRQVGFNNATANQY 151
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 692 | 89.3 | 151 | 1 | US-08-233-788A-59 |
| 2 | 514 | 66.3 | 120 | 1 | US-08-233-788A-57 |
| 3 | 92 | 11.9 | 673 | 3 | US-09-136-387-8 |
| 4 | 92 | 11.9 | 673 | 4 | US-09-841-835-8 |
| 5 | 92 | 11.9 | 949 | 3 | US-09-136-387-10 |
| 6 | 92 | 11.9 | 949 | 4 | US-09-841-835-10 |
| 7 | 92 | 11.9 | 1327 | 3 | US-09-136-387-2 |
| 8 | 92 | 11.9 | 1327 | 4 | US-09-841-835-2 |
| 9 | 92 | 11.9 | 1327 | 4 | US-09-972-115A-8 |
| 10 | 89.5 | 11.5 | 738 | 3 | US-08-864-038A-3 |
| 11 | 87 | 11.2 | 2123 | 3 | US-08-968-685A-10 |
| 12 | 83.5 | 10.8 | 339 | 4 | US-09-252-991A-32096 |
| 13 | 83.5 | 10.8 | 943 | 4 | US-09-056-556-204 |
| 14 | 83.5 | 10.8 | 943 | 4 | US-09-072-596-199 |
| 15 | 83.5 | 10.8 | 943 | 4 | US-09-477-135A-131 |
| 16 | 83.5 | 10.8 | 943 | 4 | US-09-072-967-204 |
| 17 | 82.5 | 10.6 | 2736 | 4 | US-09-252-991A-30227 |
| 18 | 81 | 10.5 | 878 | 4 | US-09-540-236-3401 |
| 19 | 80.5 | 10.4 | 873 | 4 | US-09-336-447A-13 |
| 20 | 80 | 10.3 | 892 | 4 | US-09-336-447A-5 |
| 21 | 79.5 | 10.3 | 204 | 6 | 5187262-1 |
| 22 | 79.5 | 10.3 | 518 | 3 | US-09-043-123-2 |
| 23 | 79.5 | 10.3 | 1207 | 4 | US-09-489-039A-11518 |
| 24 | 78.5 | 10.1 | 539 | 4 | US-09-719-402A-2 |
| 25 | 78.5 | 10.1 | 714 | 4 | US-09-841-786-4 |
| 26 | 78.5 | 10.1 | 1912 | 1 | US-08-409-995-4 |
| 27 | 78.5 | 10.1 | 1912 | 3 | US-08-685-467-4 |

| | | | | | | |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 78.5 | 10.1 | 2353 | 3 | US-09-377-155-33 | Sequence 33, Appl |
| 29 | 78.5 | 10.1 | 2353 | 3 | US-08-913-942-4 | Sequence 4, Appl |
| 30 | 78.5 | 10.1 | 2353 | 4 | US-09-669-974-33 | Sequence 33, Appl |
| 31 | 78.5 | 10.1 | 2353 | 4 | US-09-797-862-33 | Sequence 33, Appl |
| 32 | 78.5 | 10.1 | 2354 | 4 | US-09-268-347-47 | Sequence 47, Appl |
| 33 | 78.5 | 10.1 | 2411 | 4 | US-09-268-347-36 | Sequence 36, Appl |
| 34 | 78.5 | 10.1 | 3241 | 4 | US-09-841-786-1 | Sequence 1, Appl |
| 35 | 77.5 | 10.0 | 318 | 4 | US-09-107-532A-5818 | Sequence 5818, Ap |
| 36 | 77.5 | 10.0 | 3169 | 4 | US-09-453-702B-257 | Sequence 257, App |
| 37 | 77 | 9.9 | 415 | 4 | US-09-025-769B-280 | Sequence 280, App |
| 38 | 77 | 9.9 | 1690 | 4 | US-09-595-684B-39 | Sequence 39, Appl |
| 39 | 76.5 | 9.9 | 238 | 4 | US-09-495-880A-42 | Sequence 42, Appl |
| 40 | 76.5 | 9.9 | 956 | 4 | US-09-134-078-63 | Sequence 63, Appl |
| 41 | 76.5 | 9.9 | 1413 | 4 | US-09-252-991A-21268 | Sequence 21268, A |
| 42 | 75.5 | 9.7 | 645 | 4 | US-09-919-172-41 | Sequence 41, Appl |
| 43 | 75.5 | 9.7 | 789 | 3 | US-08-960-780-6 | Sequence 6, Appl |
| 44 | 75.5 | 9.7 | 789 | 3 | US-09-073-898-6 | Sequence 6, Appl |
| 45 | 75.5 | 9.7 | 789 | 4 | US-09-850-351A-6 | Sequence 6, Appl |

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 89.3%; Score 692; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.9e-62;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNNHGGSSGPDSTLSIYQYGSANAALAIQ 60
|||||

```

Db      1  MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY      61  SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120
Db      61  SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
QY      121  LVTRVVTHEMAHSAVMVRQVGGNNATANQY 151
Db      121  NNPALVNTQASDSSVMVRQVGGNNATANQY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 66.3%; Score 514; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.2e-44;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      22  VVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADY 81
Db      1  VVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADY 60
QY      82  GQGANSTIELTQNGFRNNATIDOWNAKNSDITVQY 118
Db      61  GQGANSTIELTQNGFRNNATIDOWNAKNSDITVQY 97

RESULT 3
US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:

```

```

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-196-387-8

Query Match 11.9%; Score 92; DB 3; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.54;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY      6  VAAFAAL-VVGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDAR 64
Db      99  VAAAPVVPVAVTSSAAGVAPNPAAGSGNNSPSSSPSTSS-SSSPSPSPGSLASPPAA 157
QY      65  KSETTIT---QSGYNGADVCGGADNSTIELTQNG--FRNNATIDOWNAKNSDI 113
Db      158  GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNGDVSRVRLVDAANVNAKM 212

RESULT 4
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-8

Query Match 11.9%; Score 92; DB 4; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.54;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGCGADNSTIETQNG--FRNATTIDQWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 5
US-09-196-387-10
Sequence 10, Application US/09196387
Patent No. 6277613
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,387
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/095,225
FILING DATE: June 10, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-387-10
Query Match 11.9%; Score 92; DB 3; Length 949;
Best Local Similarity 30.4%; Pred. No. 0.84;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGCGADNSTIETQNG--FRNATTIDQWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 6
US-09-841-835-10
Sequence 10, Application US/09841835
Patent No. 6506587
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 949 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.9%; Score 92; DB 4; Length 949;
Best Local Similarity 30.4%; Pred. No. 0.84;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;
QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQYGSANAALALQSDAR 64

```

Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGQGDADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELEACRNGDVSVRKRLVDAANVNAKDM 212

```

RESULT 7

```

US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 6217613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521

```

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-196-387-2

```

```

Query Match 11.9%; Score 92; DB 3; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

```

```

QY 6 VAAFAAI-VVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGQGDADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELEACRNGDVSVRKRLVDAANVNAKDM 212

```

RESULT 8

```

US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:

```

```

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521

```

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

```

```

Query Match 11.9%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.3;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

```

```

QY 6 VAAFAAI-VVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAAGSGNNSSPSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---OSGYNGADVGQGDADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELEACRNGDVSVRKRLVDAANVNAKDM 212

```

RESULT 9

```

US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:

```

```

; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64

```


SOFTWARE: PatentIn version 3.1

SEQ ID NO 8

LENGTH: 1327

TYPE: PRT

ORGANISM: Homo sapiens

US-09-972-115A-8

Query Match 11.9%; Score 92; DB 4; Length 1327;

Best Local Similarity 30.4%; Pred. No. 1.3;

Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDAR 64

Db 99 VAAAPVPAVSTSAAGVAENPAGSGNNPSSSSSTSS-SGSSPSPGSSLAESPEAA 157

QY 65 KSETTIT-OSQYGNAGVQGDADNSTIETLQNG--FRNNATIDOWNAKNSDI 113

Db 158 GVSSTAPLPGAGGTGVPVAGLRELLACRNGDVSRVKRLVDAANTAKDM 212

RESULT 10

US-08-864-038A-3

Sequence 3, Application US/08864038A

Patent No. 6001592

GENERAL INFORMATION:

APPLICANT: Kunio NAKASHIMA et al.

TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR

TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID

TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

TITLE OF INVENTION: TO SAID POLYPEPTIDE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: 812-5 Hirano

STREET: Ishinden

CITY: Tsu-city

STATE: Mie-prefecture

COUNTRY: JAPAN

ZIP: 514-01

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: Microsoft Windows 95

SOFTWARE: Word Perfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/864,038A

FILING DATE: May 28, 1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-184459

FILING DATE: 15-July-1996

ATTORNEY/AGENT INFORMATION:

NAME: C. Bruce Hamburg

REFERENCE/DOCKET NUMBER: F-5610

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)966-2340

TELEFAX: (212)953-7733

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 738

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Pinctada fucata

CELL TYPE: mantle epithelial cell

FEATURES:

NAME/KEY: peptide

LOCATION: from 1 to 738

IDENTIFICATION METHOD: E (by experiment)

US-08-864-038A-3

Query Match

11.5%; Score 89.5; DB 3; Length 738;

Best Local Similarity 35.4%; Pred. No. 1.1;

Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;

QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSD 62

Db 419 LLKSSASASASASASAG-----GGGGGGGGGGGGG-----GGGAGALA----- 460

QY 63 ARKSETTITQSGYNGADVGQG 84

Db 461 -----AALAAAGAGGGGGLGGGG 477

RESULT 11

US-08-968-685A-10

Sequence 10, Application US/08968685A

Patent No. 6214981

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH

APPLICANT: PLOSILA, LAURA

TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE

TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968,685A

FILING DATE: NO. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 2123 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-968-685A-10

Query Match 11.2%; Score 87; DB 3; Length 2123;

Best Local Similarity 25.4%; Pred. No. 7.6;

Matches 48; Conservative 27; Mismatches 62; Indels 52; Gaps 11;

QY 1 MKLLKVAFAAIVV-----SGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAA 56

Db 44 LSFARIAALAVLIGATNGSAYAGIGISEADGG--KGGANARG-DKSIAT--GDIAQA 97

QY 57 LALQSDARKSETTITQSGYNGADVG-QQADNSTI-----ELTQNGFRNNA--TIDOWNAK 109

Db 98 LGSQSIATGDNKIVHNS--NNNANTGAKASGNES-IAIGGDVLASGHASIAIGSDDLYLKK 155

QY 110 NSDITVGQYDQLVTRV-----VTHEMAHASVMVRQV----- 140

Db 156 E---TVQIQISELLPIIRGOKALNDIYQLADTNLQYRTHAQGHASTAGMSYAKGHPS 212

QY 141 -CFGNNATA 148

QY 12 IIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68
Db 159 IGLTGSGLGF-----GGLNSGTGN-----IGLFSGTGNVVGIGSGTGNWGICNSG 205
QY 69 TITOSGYGNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSD---ITVGOYD 119
Db 206 NSYNTGFGNSGDANTGFPFNSGIANTGVGNAGNYNTGTSYNPNGSNTGGFNMGOYN 259

RESULT 15

US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 10.8%; Score 83.5; DB 4; Length 943;
Best Local Similarity 25.4%; Pred. No. 6;
Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

QY 12 IIVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68
Db 266 IGLTGSGLGF-----GGLNSGTGN-----IGLFSGTGNVVGIGSGTGNWGICNSG 312
QY 69 TITOSGYGNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSD---ITVGOYD 119
Db 313 NSYNTGFGNSGDANTGFPFNSGIANTGVGNAGNYNTGTSYNPNGSNTGGFNMGOYN 366

Search completed: August 2, 2004, 14:58:32
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 31362936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 525 | 67.7 | 151 | 12 | US-09-741-873B-4 |
| 2 | 525 | 67.7 | 151 | 12 | US-09-741-873B-4 |
| 3 | 447 | 57.7 | 131 | 12 | US-09-741-873B-2 |
| 4 | 447 | 57.7 | 131 | 12 | US-09-741-873B-2 |
| 5 | 104 | 13.4 | 445 | 15 | US-10-369-493-20638 |
| 6 | 98 | 12.6 | 210 | 12 | US-10-425-114-53421 |
| 7 | 98 | 12.6 | 211 | 12 | US-10-425-114-53421 |
| 8 | 95 | 12.3 | 186 | 12 | US-10-282-122A-49412 |
| 9 | 92 | 11.9 | 673 | 9 | US-09-841-835-8 |
| 10 | 92 | 11.9 | 949 | 9 | US-09-841-835-10 |
| 11 | 92 | 11.9 | 1327 | 9 | US-09-841-835-2 |
| 12 | 92 | 11.9 | 1327 | 10 | US-09-972-115A-8 |
| 13 | 92 | 11.9 | 1327 | 14 | US-10-199-937-4 |
| 14 | 90.5 | 11.7 | 263 | 12 | US-10-425-114-49960 |
| 15 | 88.5 | 11.4 | 145 | 16 | US-10-437-963-147748 |

| | | | | | | |
|----|------|------|------|----|----------------------|--------------------|
| 16 | 88 | 11.4 | 369 | 12 | US-10-425-114-56041 | Sequence 56041, A |
| 17 | 88 | 11.4 | 486 | 12 | US-10-424-599-275468 | Sequence 275468, A |
| 18 | 88 | 11.4 | 507 | 12 | US-10-425-114-57763 | Sequence 57763, A |
| 19 | 87 | 11.2 | 597 | 9 | US-09-793-306-146 | Sequence 146, App |
| 20 | 87 | 11.2 | 628 | 12 | US-10-282-122A-53269 | Sequence 53269, A |
| 21 | 87 | 11.2 | 678 | 12 | US-10-282-122A-64573 | Sequence 64573, A |
| 22 | 87 | 11.2 | 2122 | 9 | US-09-813-214A-9 | Sequence 9, App.i |
| 23 | 86.5 | 11.2 | 527 | 9 | US-09-712-363-156 | Sequence 156, App |
| 24 | 86.5 | 11.2 | 527 | 14 | US-10-080-170-348 | Sequence 348, App |
| 25 | 86.5 | 11.2 | 527 | 16 | US-10-080-170-348 | Sequence 348, App |
| 26 | 86 | 11.1 | 482 | 14 | US-10-156-761-8763 | Sequence 8763, App |
| 27 | 86 | 11.1 | 486 | 15 | US-10-369-493-20619 | Sequence 20619, A |
| 28 | 85.5 | 11.0 | 270 | 16 | US-10-437-963-122263 | Sequence 122263, A |
| 29 | 85.5 | 11.0 | 1048 | 14 | US-10-174-383-56 | Sequence 56, App1 |
| 30 | 85.5 | 11.0 | 1048 | 15 | US-10-374-780A-2086 | Sequence 2086, App |
| 31 | 85 | 11.0 | 438 | 14 | US-10-156-761-9343 | Sequence 9343, App |
| 32 | 84.5 | 10.9 | 271 | 14 | US-10-156-761-11721 | Sequence 11721, A |
| 33 | 83.5 | 10.8 | 943 | 9 | US-09-996-634-131 | Sequence 131, App |
| 34 | 83.5 | 10.8 | 943 | 10 | US-09-997-182-131 | Sequence 131, App |
| 35 | 83.5 | 10.8 | 943 | 10 | US-09-997-181-131 | Sequence 131, App |
| 36 | 83.5 | 10.8 | 943 | 14 | US-10-193-002-199 | Sequence 199, App |
| 37 | 83.5 | 10.8 | 943 | 14 | US-10-084-843-204 | Sequence 204, App |
| 38 | 83.5 | 10.8 | 3300 | 12 | US-10-282-122A-64369 | Sequence 64369, A |
| 39 | 83 | 10.7 | 242 | 12 | US-10-425-114-61520 | Sequence 61520, A |
| 40 | 83 | 10.7 | 254 | 10 | US-09-880-748-1136 | Sequence 1136, App |
| 41 | 83 | 10.7 | 254 | 10 | US-09-880-748-1136 | Sequence 1136, App |
| 42 | 83 | 10.7 | 254 | 12 | US-10-293-418-1136 | Sequence 1136, App |
| 43 | 83 | 10.7 | 254 | 12 | US-10-293-418-1136 | Sequence 1136, App |
| 44 | 83 | 10.7 | 257 | 10 | US-09-880-748-1494 | Sequence 1494, App |
| 45 | 83 | 10.7 | 257 | 12 | US-10-293-418-1494 | Sequence 1494, App |

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 67.7%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVWPQWGGCGNHNHGGNSGPGDTLSIYQYGSANAALALQ 60

```
Db      1  MKLLKVAATAAIVFSSAGAVVPOYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Qy      61  SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWAKNSDITVGVQYDQ 120
Db      61  TDARNSDLTITQGGGNGADVGCGSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120
Qy      121  LVTRVVTHEMAHASVMVRQVGFGNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQFGFNATAHOY 151
```

RESULT 2

```
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1997-11-26
; PRIOR FILING DATE: 1989-05-04
; PRIOR FILING DATE: 1989-05-04
; PRIOR FILING DATE: 1991-11-06
; PRIOR FILING DATE: 1992-11-03
; PRIOR FILING DATE: 1992-11-03
; PRIOR FILING DATE: 1994-01-28
; PRIOR FILING DATE: 1994-01-28
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4
```

```
Query Match      67.7%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.1e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Qy      1  MKLLKVAATAAIVSSGALAGVVPQWGGGNGHGGGNSGDPSTLSIYQYGSANAALALQ 60
Db      1  MKLLKVAATAAIVFSSAGAVVPOYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Qy      61  SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWAKNSDITVGVQYDQ 120
Db      61  TDARNSDLTITQGGGNGADVGCGSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120
Qy      121  LVTRVVTHEMAHASVMVRQVGFGNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQFGFNATAHOY 151
```

RESULT 3

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
```

```
Qy      1  MKLLKVAATAAIVSSGALAGVVPQWGGGNGHGGGNSGDPSTLSIYQYGSANAALALQ 60
Db      1  MKLLKVAATAAIVFSSAGAVVPOYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Qy      61  SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWAKNSDITVGVQYDQ 120
Db      61  TDARNSDLTITQGGGNGADVGCGSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120
Qy      121  LVTRVVTHEMAHASVMVRQVGFGNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQFGFNATAHOY 151
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

```
Query Match      57.7%; Score 447; DB 12; Length 131;
Best Local Similarity 64.9%; Pred. No. 1.5e-37;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

Qy      21  GVVPQWGGGNGHGGGNSGDPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db      1  GVVPQYGGGNGHGGGNSGPNSELNTYQYGGNSALALQTDARNSDLTITQGGGNGAD 60
Qy      81  VQCGADNSTIELTQNGFRNNATIDQWAKNSDITVGVQYDQVTRVTHEMAHASVMVRQV 140
Db      61  VQCGSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQV 120
Qy      141  GFGNNATANOY 151
Db      121  GFGNNATAHOY 131
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

Query Match

57.7%; Score 447; DB 12; Length 131;

Best Local Similarity 64.9%; Pred. No. 1.5e-37;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVVPQGGGNGHNGSGPDSITSIYQGSANAALAQSDARKSETTITQSQVGNAGD 80
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1 GVVPQGGGNGHNGSGPNSLNTIYQGGNSALALQTDARNSDLTITCHGGNGAD 60
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 81 VGGGANSTIELTQNGFRNNATIDOWNAKNSDITVGQVDQLVTRVVTHMAHASVMYRVQV 140
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 61 VGGGSDSSIDLTRGFNLSATLDWNGKNSMTVKQFGGNGAAVDQTASNSSVNVTQV 120
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 141 GFGNATANQY 151
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 121 GFGNATANQY 131
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodopseudomonas palustris
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.4%; Score 104; DB 15; Length 445;
Best Local Similarity 26.6%; Pred. No. 0.051;
Matches 45; Conservative 25; Mismatches 61; Indels 38; Gaps 8;

QY 7 AAFPA-----IVVSGSALGVVPQGGG-----NHNGC-----GNSGGPDSTLSIQY 50
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 19 AFAFADSNTVYLQTNQGDDQANITQSGNGSVGAFTNGSGFLQENGTLUGA-NLLTVKQS 77
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 51 GSANAALQSDARKSETTITQSGYGNGADVGGCADNSTIELTQNGFRNNATIDOWNAKN 110
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 78 GNSNV-----GRDIQKQSGAGNSAAIFEGTGSDVELCQTGTSGAVPSCWNWTN 129
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

QY 111 SDITVGQVDQLVTRVVTHMAHASVM-----VRQVGFGNATANQ 150
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 130 DP---GVFNK-ITQDSSNSGSKSVIQDGKNVFISIKQGTGNSTSNQ 174
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 6
US-10-425-114-53421
; Sequence 53421, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With


```

; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match
Best Local Similarity 11.9%; Score 92; DB 9; Length 1327;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAFAAI-VYSGALAGVVPQWGGNHGNGSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGCGADNSTIELTQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 12
US-09-972-115A-8
; Sequence 8; Application US/09972115A
; Publication No. US20030032769A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 0807003C
; CURRENT APPLICATION NUMBER: US/09/972.115A
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match
Best Local Similarity 11.9%; Score 92; DB 10; Length 1327;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAFAAI-VYSGALAGVVPQWGGNHGNGSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGCGADNSTIELTQNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 13
US-10-199-937-4
; Sequence 4; Application US/10199937
; Publication No. US20030190739A1
; GENERAL INFORMATION:
; APPLICANT: Christenson, Erik
; APPLICANT: DeMaggio, Anthony J.
; APPLICANT: Goldman, Phyllis S.
; APPLICANT: Mcelligott, David L.
; TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS
; FILE REFERENCE: 27866/36559
; CURRENT APPLICATION NUMBER: US/10/199,937
; PRIOR FILING DATE: 2002-07-22
; CURRENT APPLICATION NUMBER: US/09/606,035

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-14
Perfect score: 775
Sequence: 1 MKLKVAFAAIVVSGSALA.....HASYMVRQVGFGNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US160_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
|------------|-------|-------|--------|-------|-------------|

| | | | | | | |
|----|-----|-------|-----|----|----------------------|-------------------|
| 1 | 775 | 100.0 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 2 | 714 | 92.1 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 3 | 696 | 89.8 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 4 | 693 | 89.4 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 5 | 692 | 89.3 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 6 | 659 | 85.0 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 7 | 612 | 79.0 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 8 | 611 | 78.8 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 9 | 609 | 78.6 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 10 | 606 | 78.2 | 131 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 11 | 603 | 77.8 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 12 | 602 | 77.7 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 13 | 578 | 74.6 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 528 | 68.1 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 525 | 67.7 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 525 | 67.7 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 523 | 67.5 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 523 | 67.5 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 514 | 66.3 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 474 | 61.2 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 21 | 472 | 60.9 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 472 | 60.9 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 23 | 472 | 60.9 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 24 | 447 | 57.7 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 447 | 57.7 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 347 | 44.8 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 27 | 343 | 44.3 | 109 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 28 | 243 | 31.4 | 70 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 29 | 237 | 30.6 | 48 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 30 | 115 | 14.9 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 31 | 115 | 14.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 115 | 14.9 | 186 | 30 | US-10-417-886-5833 | Sequence 5833, Ap |
| 33 | 104 | 13.5 | 151 | 19 | US-09-543-407-8 | Sequence 8, Appl |
| 34 | 104 | 13.4 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 35 | 104 | 13.4 | 445 | 33 | US-60-360-039-20638 | Sequence 20638, A |
| 36 | 98 | 12.7 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 37 | 98 | 12.6 | 210 | 28 | US-10-219-999-43013 | Sequence 43013, A |
| 38 | 98 | 12.6 | 210 | 30 | US-10-425-114-53421 | Sequence 53421, A |
| 39 | 98 | 12.6 | 210 | 30 | US-10-425-114A-53421 | Sequence 53421, A |
| 40 | 98 | 12.6 | 210 | 33 | US-60-312-544-6765 | Sequence 6765, Ap |
| 41 | 98 | 12.6 | 210 | 33 | US-60-324-109-30520 | Sequence 30520, A |
| 42 | 98 | 12.6 | 210 | 30 | US-10-424-599-245046 | Sequence 245046, |
| 43 | 95 | 12.3 | 186 | 28 | US-10-282-122A-49412 | Sequence 49412, A |
| 44 | 92 | 11.9 | 673 | 20 | US-09-678-411-8 | Sequence 8, Appl |
| 45 | 92 | 11.9 | 949 | 20 | US-09-678-411-10 | Sequence 10, Appl |

ALIGNMENTS

RESULT 1
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afaA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.

US-09-543-407-14

| | Query Match | 100.0% | Score 775; | DB 19; | Length 151; |
|----|-----------------------|--|--------------------|-----------|-------------|
| | Best Local Similarity | 100.0%; | Pred. No. 6.1e-75; | | |
| | Matches 151; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | MKLLKVAFAAIVVSGSALAGVPGWGGGHNHGGNGSGPDSLSIYQYGSANAALAQ | 60 | | |
| Db | 1 | MKLLKVAFAAIVVSGSALAGVPGWGGGHNHGGNGSGPDSLSIYQYGSANAALAQ | 60 | | |
| Qy | 61 | SDARKSETITQSGYNGADVGQAGDNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ | 120 | | |
| Db | 61 | SDARKSETITQSGYNGADVGQAGDNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ | 120 | | |
| Qy | 121 | LVTRVVTHEMAHASVMVRQVGFGNNTANQY | 151 | | |
| Db | 121 | LVTRVVTHEMAHASVMVRQVGFGNNTANQY | 151 | | |

RESULT 2:

```

US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FEMBRAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PRT3 of Leishmania major.
US-09-543-407-24

```

| | | | | |
|-----------------------|-----------------|--|------------|-------------|
| Query Match | 92.1% | Score 714; | DB 19; | Length 151; |
| Best Local Similarity | 91.1% | Prod. No. 2.4e-68; | | |
| Matches 144; | Conservative 0; | Mismatches 0; | Indels 14; | Gaps 2; |
| QY | 1 | MKLLKVAFAAIVVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIIQYGSANAALQ | 60 | |
| DBb | 1 | MKLLKVAFAAIVVSGSALAGVVPQGGGNNHNGGNSGPDSTLSIIQYGSANAALQ | 60 | |
| QY | 61 | SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDOWNAKNSDIITVGQYDQ | 120 | |
| DBb | 61 | SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDOWNAKN-----YDQ | 113 | |
| QY | 121 | LVTRVVTHEMAHA-----SVNVROVGFGNNTANQY | 151 | |
| DBb | 114 | LVTRVVTHEMAHANCTASDSVMNVROVGFGNNTANQY | 151 | |

RESULT 3

```

RES001.3
/ Sequence 12, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIBRILAR SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407

```

```

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-03-543-407-12

Query Match 89.8%; Score 696; DB 19; Length 151;
Best Local Similarity 87.6%; Pred. No. 2.1e-66;
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

Qy 1 MKLLKVAFAAAIVVSGSALAGVFPQWGGGHNHGGNGSSGPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAAIVVSGSALAGVFPQWGGGHNHGGNGSSGPDSTLSIYQYGSANAALQ 60
Qy 61 SPARKSETTITCSGYNGADVGGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGG--- 117
Db 61 SPARKSETTITCSGYNGADVGGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGG 120
Qy 118 -----YDQLVTRVTVTHMAHASVMVRQVGFGNATANQY 151
Db 121 NNAALVNYDQLVTRVTVTHMAHA-----NNATANQY 151

```

RESULT 4

```

US-09-543-407-5
RESOLUT 4
/ Sequence S, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Doran, James L.
/ APPLICANT: Collinsen, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 5
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Salmonella enteritidis
US-09-543-407-5

```

| | Query Match | 89.4% | Score 693; | DB 19; | Length 151; |
|----|-----------------------|--|--------------------|--------|-------------|
| | Best Local Similarity | 90.7%; | Prod. No. 4.4e-66; | | |
| | Matches 137; | Conservative | 3; Mismatches 11; | Indels | 0; Gaps |
| Qy | 1 | MKLLKVAAFRAIIVVSGSALAGVVPQGGGNGNNGGNSGGPDSTLUSIIQYGSANAALALQ | 60 | | |
| Db | 1 | MKLLKVAAFRAIIVVSGSALAGVVPQGGGNGNNGGNSGGPDSTLUSIIQYGSANAALALQ | 60 | | |
| Qy | 61 | SDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNNATIDOWNAKNSDITVGQYDQ | 120 | | |
| Db | 61 | SDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNNATIDOWNAKNSDITVGQYGG | 120 | | |
| Qy | 121 | LTVRVVTHEMAHASVMVTVQVGFNNATANQY | 151 | | |
| Db | 121 | NNAALVNQATSDSSVMVTVQVGFNNATANQY | 151 | | |

RESULT 5

US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDBANBERY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-57

Query Match      89.3%; Score 692; DB 6; Length 151;
Best Local Similarity 90.7%; Pred. No. 5,6e-66;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGNFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGNFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
Db 121 NNPAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26
Query Match      79.0%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.5e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGNFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGNFRNNATIDQWNAKNSDITVGYDQ 120
QY 121 LVTRVVTHEMAHASVVRQVGFNNATANQY 151
Db 121 NNPAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match          78.8%; Score 611; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.2e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SPARKSETTITSGYNGADVCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SPARKSETTITSGYNGADVCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 LVTRVVTHEMAHASVVMVRQVFGNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 9
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match          78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.2e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITSGYNGADVCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYDQ 120
DB 61 SPARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-31

Query Match          78.2%; Score 606; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 9e-57;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60

QY 81 VCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYDQLVTRVVTHEMAHASVVMVRQV 140
DB 61 VCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYDQLVTRVVTHEMAHASVVMVRQV 120

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
; US-09-543-407-31

Query Match          78.2%; Score 606; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 9e-57;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60

QY 81 VCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYDQLVTRVVTHEMAHASVVMVRQV 140
DB 61 VCGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYDQLVTRVVTHEMAHASVVMVRQV 120

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match          77.8%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYGNAGADVQDLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 LVTRVVTHEMAHASVVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 77.7%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3e-56; Indels 0; Gaps 0;
Matches 123; Conservative 5; Mismatches 23;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYGNAGADVQDLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 LVTRVVTHEMAHASVVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.6%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQVLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYGNAGADVQDLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 120
QY 121 LVTRVVTHEMAHASVVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.1%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.9e-48; Indels 0; Gaps 0;
Matches 104; Conservative 20; Mismatches 27;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYGNAGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 TDARNSDLTITQGGGNGADVQGGSDSSIDLITQRFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 LVTRVVTHEMAHASVVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      67.7%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 6.1e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQVGGGNGHGGNNGSGDPSTLSIYQGSANAALALQ 60
DB      1 MKLLKVAATAAIVFSGSAGVVPQVGGGNGHGGNNGSGNPNSELNIYQGGNSALALQ 60

QY      61 SPARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYDQ 120
DB      61 TDARNSDLTITQHGCGNGADVGCGSDSDSIDLTQRFNGNSATLDQNGKNSMTVKQFEG 120

QY      121 LVTRVVTHEMAHASVMVROVGFNNATANQY 151
DB      121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

Search completed: August 2, 2004, 15:26:42
Job time : 167.9 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-14
Perfect score: 775
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....HASVMRVQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New:*
1: /cgn2_6/ptodata/2/paa/PCT NEW COMB.pbp:*
2: /cgn2_6/ptodata/2/paa/US06 NEW COMB.pbp:*
3: /cgn2_6/ptodata/2/paa/US07 NEW COMB.pbp:*
4: /cgn2_6/ptodata/2/paa/US08 NEW COMB.pbp:*
5: /cgn2_6/ptodata/2/paa/US09 NEW COMB.pbp:*
6: /cgn2_6/ptodata/2/paa/US10 NEW COMB.pbp:*
7: /cgn2_6/ptodata/2/paa/US60 NEW COMB.pbp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----------------------|--------------------|
| 1 | 525 | 67.7 | 151 | US-09-741-873C-4 | Sequence 4, Appli |
| 2 | 447 | 57.7 | 131 | US-09-741-873C-2 | Sequence 2, Appli |
| 3 | 92 | 11.9 | 299 | US-10-170-205E-35751 | Sequence 35751, A |
| 4 | 92 | 11.9 | 1203 | US-10-170-205E-741 | Sequence 741, Appl |
| 5 | 92 | 11.9 | 1327 | PCT-US04-02338-49 | Sequence 49, Appl |
| 6 | 90.5 | 11.7 | 258 | US-10-425-115-300390 | Sequence 300390, |
| 7 | 89.5 | 11.5 | 256 | US-10-425-115-301334 | Sequence 301334, |
| 8 | 89 | 11.5 | 295 | US-10-425-115-312468 | Sequence 312468, |
| 9 | 88 | 11.4 | 520 | US-10-479-638-21 | Sequence 21, Appl |
| 10 | 86.5 | 11.2 | 527 | US-10-468-356-348 | Sequence 348, Appl |
| 11 | 86 | 11.1 | 179 | US-10-425-115-346132 | Sequence 346132, |
| 12 | 85.5 | 11.0 | 1048 | PCT-US04-05654-2086 | Sequence 2086, Ap |
| 13 | 82.5 | 10.6 | 535 | PCT-US04-21492-88 | Sequence 88, Appl |
| 14 | 82 | 10.6 | 758 | US-60-565-632-9194 | Sequence 9194, Ap |
| 15 | 82 | 10.6 | 758 | US-60-579-082-9194 | Sequence 9194, Ap |
| 16 | 82 | 10.6 | 1095 | PCT-US03-24982A-273 | Sequence 273, Appl |
| 17 | 82 | 10.6 | 2228 | US-10-745-237-96 | Sequence 96, Appl |
| 18 | 81.5 | 10.5 | 376 | US-10-491-733-2 | Sequence 2, Appli |
| 19 | 81.5 | 10.5 | 508 | US-10-425-115-285216 | Sequence 285216, |
| 20 | 81.5 | 10.5 | 532 | US-10-425-115-285214 | Sequence 285214, |
| 21 | 81 | 10.5 | 131 | US-60-565-632-11109 | Sequence 11109, A |
| 22 | 81 | 10.5 | 131 | US-60-579-062-11109 | Sequence 11109, A |
| 23 | 81 | 10.5 | 201 | US-10-425-115-309662 | Sequence 309662, |
| 24 | 81 | 10.5 | 376 | US-10-425-115-342526 | Sequence 342526, |
| 25 | 80.5 | 10.4 | 246 | US-10-854-439-511 | Sequence 511, App |
| 26 | 80.5 | 10.4 | 388 | US-09-248-796A-17306 | Sequence 17306, A |

ALIGNMENTS

RESULT 1

US-09-741-873C-4

: Sequence 4; Application US/09741873C

: GENERAL INFORMATION:

: APPLICANT: Normark, Staffan

: APPLICANT: Olsen, Arne

: TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

: FILE REFERENCE: 012889-084

: CURRENT APPLICATION NUMBER: US/09/741,873C

: CURRENT FILING DATE: 2000-12-22

: PRIOR APPLICATION NUMBER: SE 8801723-1

: PRIOR FILING DATE: 1998-05-06

: PRIOR APPLICATION NUMBER: US 08/978,878

: PRIOR FILING DATE: 1997-11-26

: PRIOR APPLICATION NUMBER: US 07/347,189

: PRIOR FILING DATE: 1989-05-04

: PRIOR APPLICATION NUMBER: US 07/789,437

: PRIOR FILING DATE: 1991-11-06

: PRIOR APPLICATION NUMBER: US 07/970,846

: PRIOR FILING DATE: 1992-11-03

: PRIOR APPLICATION NUMBER: US 08/187,865

: PRIOR FILING DATE: 1994-01-28

: PRIOR APPLICATION NUMBER: US 08/318,519

: PRIOR FILING DATE: 1994-10-05

: NUMBER OF SEQ ID NOS: 11

: SOFTWARE: PatentIn version 3.0

: SEQ ID NO 4

: LENGTH: 151

: TYPE: PRT

: ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match

Best Local Similarity 67.7%; Score 525; DB 5; Length 151;

Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNNGGNSGSPDSTLSIYQVGSANAALQ | 60 |
| Db | 1 | MKLLKVAATAAIVFSSAVAGVVPQYGGGNGGGGNNSGPSELNIYQVGGNSALALQ | 60 |
| Qy | 61 | SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNAIDQNAKNSDITVQYDQ | 120 |
| Db | 61 | TDARNSDLATITQGGGNGADVGQGDSDSIDLTQRFNGSATLDQNGKNSMTVKQFG | 120 |
| Qy | 121 | LVTRVTVTHEMAHASVMVRQVFGNNATANQY | 151 |
| Db | 121 | NGGAVDQDTASNSVNTVQVFGNNATANQY | 151 |

```
RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsken, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match 57.7%; Score 447; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 1.2e-31;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVVPPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 80
DB 1 GVVPPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 60
QY 81 VGGGADNSTIETONGFRNNATIDOWNAKNSDITGVQYDQLVTRVVTHEMAHASVMVQV 140
DB 61 VGGGSDSSIDITQRFNGSAILDQNGKNSEMTVKQFGGNGAAVDQTASNVVTQV 120
QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

RESULT 3
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

Query Match 11.9%; Score 92; DB 6; Length 299;
Best Local Similarity 30.4%; Pred. No. 2.2;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
```

```
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSXKRLVDAANVNAKDM 212

RESULT 4
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

Query Match 11.9%; Score 92; DB 6; Length 1203;
Best Local Similarity 30.4%; Pred. No. 11;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVGQADNSTIETLQNG--FRNNATIDOWNAKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLEACRNGDVSXKRLVDAANVNAKDM 212

RESULT 5
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: USE
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match 11.9%; Score 92; DB 1; Length 1327;
Best Local Similarity 30.4%; Pred. No. 12;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSPSSSPTSS-SSSSPSPGSLAESPEAA 157
```

QY 65 KSETTIT---QSGYNGADYQGADNSTIELTQNG--FRNATIDQWNAKNSDI 113
Db 158 GVSSITAPLPGAGGTGVPVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

RESULT 6
US-10-425-115-300390
; Sequence 300390, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300390
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1):(258)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_37025C.1.pep
US-10-425-115-300390

Query Match 11.7%; Score 90.5; DB 6; Length 258;
Best Local Similarity 27.9%; Pred. No. 2.5;
Matches 29; Conservative 15; Mismatches 39; Indels 21; Gaps 4;

QY 17 SALAGVFPQWGGGNGHNGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSG-- 74
Db 78 SSIAG-----GGGGGQGGGTNGSGSGGGGSGGSGSTAAAGPSSGNYDAEGRGAG 132

QY 75 -----YNGA--DVQGANSTIEL--TONGFRNATID 104
Db 133 GGMGGGANGAYGSGAGGKGEGVSGVALAPSSNGYNGGAAD 176

RESULT 7
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1):(256)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pep
US-10-425-115-301334

Query Match 11.5%; Score 89.5; DB 6; Length 256;
Best Local Similarity 27.9%; Pred. No. 3;
Matches 29; Conservative 14; Mismatches 40; Indels 21; Gaps 4;

QY 17 SALAGVFPQWGGGNGHNGSSGPDSTLSIYQGSANAALALQSDA---RKSETTITQ 72
Db 158 GVSSITAPLPGAGGTGVPVSGALRELLEACRNGDVSRVKRLVDAANVNAKDM 212

Db 79 SSVAG-----GGGGGQGGGTNGSGSGGGSGTSTTAASGPFSSGNYANAEKGAG 133
QY 73 SGYNGAD-----VGQGANSTIEL--TONGFRNATID 104
Db 134 GGMGGGADGAYGAGGKGQGBSGVALAPSSDGYNGGAAD 177

RESULT 8
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1):(295)
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-425-115-312468

Query Match 11.5%; Score 89; DB 6; Length 295;
Best Local Similarity 35.1%; Pred. No. 3.9;
Matches 33; Conservative 10; Mismatches 39; Indels 12; Gaps 4;

QY 26 WGGGNGHNGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADYQGA 85
Db 138 YGGGYSYSSGYSYG--GYAANGYGVSGSGNYNAGSGGYSGS----DGYNGAASGGYA 192

QY 86 DNSTIELTQNGFRNATIDQWNAKNSDITVGYD 119
Db 193 NNLSGYSNGRYN--TIG-----SSDGTGGYN 219

RESULT 9
US-10-479-638-21
; Sequence 21, Application US/10479638
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: WYO.02-0004US
; CURRENT APPLICATION NUMBER: US/10/479,638
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/US02/18256
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Argiope trifasciata
US-10-479-638-21

Query Match 11.4%; Score 88; DB 6; Length 520;
Best Local Similarity 21.2%; Pred. No. 9.2;
Matches 35; Conservative 30; Mismatches 66; Indels 38; Gaps 6;

QY 15 SGSALAGVFPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSET-TITQS 73
Db 347 AGAAAAAASAGAGAGGCGGCGVAGGSS-----ISYCATSSSATSSSTASSRSRGTG 402

```
QY 74 GYNGADVGGADNSTIELTONG-FRNATIDOWNAKNSDI----- 113
DB 403 GYGAGAAAGAGAAAGAGSGSISRLSSABAVRVSNSICAVASGASALPGVISNIF 462
QY 114 -----TVGYDQDLVTR-----VVTHEMAHASV-MVRQVGFNNATNQ 150
DB 463 SGVSSAGSYEEAVIQSLLEVLVLLHILSNSSIGYVGAEGUGDSIAVVQ 512

RESULT 10
US-10-468-356-348
; Sequence 348, Application US/10468356
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENEOMICS AS A TOOL FOR
; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 05394.0019
; CURRENT APPLICATION NUMBER: US/10/468,356
; PRIOR FILING DATE: 2003-08-19
; PRIOR APPLICATION NUMBER: 10/080,170
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 655
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 348
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-468-356-348

Query Match 11.2%; Score 86.5; DB 6; Length 527;
Best Local Similarity 22.9%; Pred. No. 13;
Matches 35; Conservative 13; Mismatches 56; Indels 49; Gaps 5;

QY 21 GVVPOWGG---CGNHNGGSSGPD-----STLSIYQYGSANAALALQSPARKSE 67
DB 387 GVAPOGGVYAEAPAGRDYDYGSGADPDYQAPAGSYGYGQYSGAGTSVTLQDDGSGR 446
QY 68 TTITSG---YNGADV-----GQADNSTIELTQNGFRNNATIDQ 105
DB 447 TQLREGSNIKRGQDAQPLDPTGVSRRLHBIWDGQVALLADLNSGTGTVNNAVQOE 506
QY 106 WNAKNSDIYQYDQLVTRVTHEMAHASVMVR 138
DB 507 WQADGDVT-----RLGHSBIIVR 525

RESULT 11
US-10-425-115-346132
; Sequence 346132, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; the Development of Plants
; FILE REFERENCE: 38-21(S3222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346132
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pap
US-10-425-115-346132

Query Match 11.1%; Score 86; DB 6; Length 179;
Best Local Similarity 29.9%; Pred. No. 4.1;

QY 19 LAGVVPWGGGG-----NHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTI 70
DB 64 LDLLGLGGGLGDLGGTGGGKKNKAQDSGNAEGSNA-----QEDSGNAQ--- 116
QY 71 TQSGYNGADVGQCADNSTIELTQNGFRNNATIDOWNAKNSDIYVQ 117
DB 117 EESGNQACAGAGA-----ENGAANNGT-EAAGAENAAAGNSQ 154

Matches 32; Conservative 11; Mismatches 40; Indels 24; Gaps 5;

QY 19 LAGVVPWGGGG-----NHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTI 70
DB 64 LDLLGLGGGLGDLGGTGGGKKNKAQDSGNAEGSNA-----QEDSGNAQ--- 116
QY 71 TQSGYNGADVGQCADNSTIELTQNGFRNNATIDOWNAKNSDIYVQ 117
DB 117 EESGNQACAGAGA-----ENGAANNGT-EAAGAENAAAGNSQ 154

RESULT 12
PCT-US04-05654-2086
; Sequence 2086, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omaira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttersen, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E
; APPLICANT: Kumimoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 2086
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1152 Paralogous to G1146
PCT-US04-05654-2086

Query Match 11.0%; Score 85.5; DB 1; Length 1048;
Best Local Similarity 25.7%; Pred. No. 34;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GGGGNGGSSGPDSTLSI-----YQYGSANAALALQSDARKSETTITQSGYG 76
DB 95 GGYGGGGGSGSPGPPQPSVPELHOATSTTYQAVSSQTLSEVSPQVPEPTVLAQQF- 153
QY 77 NGADVCGQADNSTIE-----LTQNGFRNNATIDOWNAKNSDIYVQYDQ 120
DB 154 EQLSVEQGFAPSAIQIPSSSKAFKPFMPRPGKQKRCIVKANHFPAELPKDLHLYDV 213
QY 121 LVTRVVTHEMAHASVMVRV 140
DB 214 TIPEVTSRGVNRVNMQLV 233

RESULT 13
PCT-US04-21492-88
; Sequence 88, Application PC/TUS0421492
; GENERAL INFORMATION:
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-14
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVGFNNATANQY 151
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|----------|--------------------|
| 1 | 693 | 89.4 | 151 | 2 JC6039 | fimbrin protein ag |
| 2 | 693 | 89.4 | 151 | 2 A10635 | major curlin chain |
| 3 | 528 | 68.1 | 151 | 2 S70788 | curlin protein csg |
| 4 | 506.5 | 65.4 | 152 | 2 D50806 | curlin major subun |
| 5 | 506.5 | 65.4 | 152 | 2 H85665 | hypothetical prote |
| 6 | 104.5 | 13.5 | 151 | 2 S70787 | curlin nucleator p |
| 7 | 104.5 | 13.5 | 151 | 2 C90806 | minor curlin subun |
| 8 | 104.5 | 13.5 | 151 | 2 G85665 | curlin minor chain |
| 9 | 98.5 | 12.7 | 151 | 2 JC6040 | fimbrin protein ag |
| 10 | 96.5 | 12.5 | 552 | 2 D70604 | probable PPE prote |
| 11 | 96 | 12.4 | 770 | 2 T51024 | related to C2H2 zi |
| 12 | 95 | 12.3 | 1748 | 2 S42136 | cnjB protein - Tet |
| 13 | 93 | 12.0 | 2174 | 2 E95965 | hypothetical glyci |
| 14 | 91 | 11.7 | 301 | 2 B84533 | hypothetical prote |
| 15 | 91 | 11.7 | 1028 | 2 A56038 | DNA-binding protei |
| 16 | 91 | 11.7 | 1213 | 2 S26356 | ovo protein - fru1 |
| 17 | 90.5 | 11.7 | 145 | 2 A3143 | conserved hypothet |
| 18 | 90.5 | 11.7 | 145 | 2 H98144 | hypothetical prote |
| 19 | 90.5 | 11.7 | 151 | 2 AH0635 | nucleation compone |
| 20 | 90 | 11.6 | 479 | 2 JN0891 | metalloproteinase |
| 21 | 89.5 | 11.5 | 256 | 2 T03371 | glycine-rich prote |
| 22 | 89.5 | 11.5 | 262 | 2 S02075 | tail fiber protein |
| 23 | 89 | 11.5 | 573 | 2 C86266 | F319.21 protein - |
| 24 | 88.5 | 11.4 | 479 | 2 A38307 | metalloproteinase |
| 25 | 87 | 11.2 | 590 | 2 E70946 | probable PPE prote |
| 26 | 87 | 11.2 | 646 | 1 S19916 | leishmanolysin (EC |
| 27 | 87 | 11.2 | 678 | 2 A70762 | probable PPE prote |
| 28 | 86.5 | 11.2 | 478 | 2 JN0892 | metalloproteinase |
| 29 | 86.5 | 11.2 | 527 | 2 B70700 | hypothetical prote |

| | | | | | |
|----|------|------|------|----------|--------------------|
| 30 | 86.5 | 11.2 | 645 | 2 F70825 | probable PPE prote |
| 31 | 86 | 11.1 | 447 | 2 G84687 | probable disease r |
| 32 | 86 | 11.1 | 599 | 2 B42049 | leishmanolysin (EC |
| 33 | 86 | 11.1 | 599 | 2 A44951 | leishmanolysin (EC |
| 34 | 85.5 | 11.0 | 575 | 2 S35327 | protein kinase sgg |
| 35 | 85 | 11.0 | 347 | 2 B39112 | merozoite 45K surf |
| 36 | 85 | 11.0 | 967 | 2 S68852 | hypothetical prote |
| 37 | 84.5 | 10.9 | 440 | 2 AD1539 | probable sugar ABC |
| 38 | 84 | 10.8 | 407 | 2 T21956 | hypothetical prote |
| 39 | 84 | 10.8 | 590 | 1 A45621 | leishmanolysin (EC |
| 40 | 84 | 10.8 | 602 | 1 PL0221 | leishmanolysin (EC |
| 41 | 83.5 | 10.8 | 285 | 2 H84219 | hypothetical prote |
| 42 | 83.5 | 10.8 | 340 | 2 A83401 | hypothetical prote |
| 43 | 83.5 | 10.8 | 440 | 2 T15352 | hypothetical prote |
| 44 | 83.5 | 10.8 | 1067 | 2 S35423 | protein kinase sgg |
| 45 | 83.5 | 10.8 | 3300 | 2 D70575 | probable PPE prote |

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6039; PC6015; A44898
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfABC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6039
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43599.1; PID:gl184714
A:Accession: PC6015
A:Molecule type: protein
A:Residues: 21-52 <CO2>
A:Experimental source: strain 27655-3b
A>Note: the authors translated the codon ACG for residue 44 as Ile
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella
A:Reference number: A44898; MUID:91310586; PMID:1677357

A:Contents: 27655
A:Accession: A44898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 21-33 <CO3>
A>Note: sequence extracted from NCBI backbone (NCBIP:45936)
C:Genetics:
A:Gene: agfa
C:Function:
A:Description: major component of thin aggregative fimbriae
A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fimbrin protein agfa #status experimental <MAT>

| | | | | | | | |
|-----------------------|-------|--|-------|------------|-----|--------|-----|
| Query Match | 89.4% | Score | 693 | DB | 2 | Length | 151 |
| Best Local Similarity | 90.7% | Pred. No. | 58-52 | | | | |
| Matches | 137 | Conservative | 3 | Mismatches | 11 | Indels | 0 |
| Gaps | 0 | | | | | | |
| Qy | 1 | MKLLKVAFAAIVVSGSALA | 1 | Qy | 60 | | |
| Db | 1 | MKLLKVAFAAIVVSGSALA | 1 | Db | 60 | | |
| Qy | 61 | SDARKSETTITGSGYNGADVCGGADNSITELTQNGFRNNATIDQWNAKNSDITVGYDQ | 120 | Qy | 120 | | |
| Db | 61 | SDARKSETTITGSGYNGADVCGGADNSITELTQNGFRNNATIDQWNAKNSDITVGYDQ | 120 | Db | 120 | | |
| Qy | 121 | LTVRVVTHEMAHSMVMVRQVGFNNATANQY | 151 | Qy | 151 | | |
| Db | 121 | LTVRVVTHEMAHSMVMVRQVGFNNATANQY | 151 | Db | 151 | | |

Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635
C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
A:Note: this species has also been called Salmonella typhimurium
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 89.4%; Score 693; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 5e-52;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGPDPSTLSIYYGSAANAALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGPDPSTLSIYYGSAANAALQ 60
QY 61 SPARKSETTITGSGYNGADVGADGADNSTIETQNGFRNATIDOWNAKNSDITVGOYDQ 120
DB 61 SPARKSETTITGSGYNGADVGADGADNSTIETQNGFRNATIDOWNAKNSDITVGOYDQ 120
QY 121 LVTRVVTHMAHASVMVRQVGFNNATANQY 151
DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788
C:Title: curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and C:Reference number: S70788; MUID:96414468; PMID:8817489
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAAG2282.1; PID:g1147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co: A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:197426617; PMID:9278503
A:Accession: G64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279; A:Experimental source: strain K-12, substrain MG1655
R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6,'V', 8-151 <OLG1>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42;44-50 <OLS2>
R:Olsen, A.N.; Arngqvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133, 'PORDSGWLW' <OLS3>
A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tr: A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 68.1%; Score 528; DB 2; Length 151;
Best Local Similarity 68.9%; Pred. No. 5.4e-38;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGPDPSTLSIYYGSAANAALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGPDPSTLSIYYGSAANAALQ 60
QY 61 SPARKSETTITGSGYNGADVGADGADNSTIETQNGFRNATIDOWNAKNSDITVGOYDQ 120
DB 61 SPARKSETTITGSGYNGADVGADGADNSTIETQNGFRNATIDOWNAKNSDITVGOYDQ 120
QY 121 LVTRVVTHMAHASVMVRQVGFNNATANQY 151
DB 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 4

D90806
C:Title: curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G: A.; Saito, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen: A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 3.6e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQW-GGGGNGGNGSGPDPSTLSIYYGSAANAALQ 59
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGPDPSTLSIYYGSAANAALQ 60
QY 60 QSDARKSETTITGSGYNGADVGADGADNSTIETQNGFRNATIDOWNAKNSDITVGOYD 119

Db 61 QADARNSLTTTQHGGNGADVGQSDSSIDLTRQFGNSATLDQWNGKDSHMTVXQFG 120
QY 120 QLVTRVTHMAHASVVMRVQVFGNNATAHQY 151
Db 121 GGNGAANDQTASNSTVNTVQVFGNNATAHQY 152
RESULT 5
H85665
Hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AB005174; NID:g12514574; PIDN:AA055788.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA
Query Match 65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 3.6e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
QY 1 MKLLKVAFAIVVSGSALAGVWPQW-CGGNGHNGGNSGPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAFAIVVSGSALAGVWPQYGGGNGHNGGNSGPNSELNLYQYGGNSALAL 60
QY 60 QSDARKSTTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQD 119
Db 61 QADARNSLTTTQHGGNGADVGQSDSSIDLTRQFGNSATLDQWNGKDSHMTVXQFG 120
QY 120 QLVTRVTHMAHASVVMRVQVFGNNATAHQY 151
Db 121 GGNGAANDQTASNSTVNTVQVFGNNATAHQY 152
RESULT 6
S70787
curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and Co
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70787
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HMS>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563
A:Experimental source: strain K12, substrain W3110
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A6720; MUID:197426617; PMID:9278503
A:Accession: F64846
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA074125.1; PID:g1787278;

A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: csgB
A:Map position: 23.15
C:Function:
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: minor curlin chain #status predicted <MAT>
Query Match 13.5%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.054;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;
QY 9 FFAIVVSGSAL--AGVVPWQGGNGHNGGNSGPDSTLSIY-QYGSANAALALQSDARK 65
Db 33 FAVNELSKSFNQAAII--GQAGTNNSAQLRQGGSKLLAVVAQEGSSNRA-KIDQTGDY 88
QY 66 SETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQDQLVTRV 125
Db 89 NLAYIDQASANDASISQAGYGNNTAMIIQKSGNKNANITQYGTQKTAIVVQSQSWAIRV 148
RESULT 7
C90806
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subs
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90806
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:g13360879; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: Ecs1419
Query Match 13.5%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.054;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;
QY 9 FFAIVVSGSAL--AGVVPWQGGNGHNGGNSGPDSTLSIY-QYGSANAALALQSDARK 65
Db 33 FAVNELSKSFNQAAII--GQAGTNNSAQLRQGGSKLLAVVAQEGSSNRA-KIDQTGDY 88
QY 66 SETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQDQLVTRV 125
Db 89 NLAYIDQASANDASISQAGYGNNTAMIIQKSGNKNANITQYGTQKTAIVVQSQSWAIRV 148
RESULT 8
G85665
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AA055787.1; GSPDB:GN00145; UWGP:Z1

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: cs9B

Query Match 13.5%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.054;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FRAIVVSGAL--AGVVPWGGGNGHNGSSGPDSTLSIY-OYGSANAALALQSDARK 65
DB 33 FAVNELSKSFFNOAII---GAGTNSAQLFGGSKLLAVVAQEGSSNRA-KIDQTGDY 88
QY 66 SETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQLVTRV 125
DB 89 NLAIDQASANDASISQAGYNTAMIIQKSGENKANIYGTQKTAIVVQPSQMAIRV 148

RESULT 9

JC6040

fimbrin protein agfB precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: JC6040

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A:Reference number: JC6039; MUID:96146512; PMID:8550497

A:Accession: JC6040

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:gl184712; PIDN:ARC43598.1; PID:gl184713

A:Experimental source: strain 276755-3b

C:Genetics:

A:Gene: agfB

C:Function:

A:Description: minor component of thin aggregative fimbriae

A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: fimbrin protein agfB #status predicted <WAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;
Best Local Similarity 26.2%; Pred. No. 0.17;
Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;

QY 10 ARAIVVSGALAGVVPWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDARKSETT 69
DB 58 ARVRQEGSKLLSVISO--EGGNRAKVDQAGNYNFAYIEQTGNAN-----DAS 103

QY 70 ITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQLVTRVTHE 129
DB 104 IQSAYGNSA-----AIIQKSGENKANIYGTQK-----TAVVQK 140

QY 130 MAHASVMVRQ 139

DB 141 QSHMAIRVQ 150

RESULT 10

D70604

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: D70604

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

i Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70604

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-552 <COL>

A:Cross-references: GB:292774; GB:AL123456; NID:g3261729; PIDN:CAB07133.1; PID:gl877289

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

C:Superfamily: Phasesolus glycine-rich cell wall protein 1.8

Query Match 12.5%; Score 96.5; DB 2; Length 552;
Best Local Similarity 28.5%; Pred. No. 1.1;
Matches 41; Conservative 10; Mismatches 62; Indels 31; Gaps 6;

QY 15 SGLSALGV-----VPQWGGGNGHNGSSGPD---STLSIYQYGSANAALALQSDA 63
DB 353 SSGNIGVFNTGANTLVP---GDLNLLGVGNSGNANIGFNGAGVLTNTFGNASILNTGLG 409

QY 64 RKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQLV 123
DB 410 NAGE---LNTGFGNAGFVNTGFDNSGNVNTGNSGNINTGWNAGNVNTGFG----- 459

QY 124 RVVTHEMAHASVMVRQVGFQGNAT 147

DB 460 -IITDSGLTNS-----GFGNTCT 476

RESULT 11

T51024

N:Alternate names: protein B7F21.50

C:Species: Neurospora crassa

C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000

C:Accession: T51024

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura

submitted to the Protein Sequence Database, July 2000

A:Reference number: Z25286

A:Accession: T51024

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-770 <SCH>

A:Cross-references: EMBL:AJ389901; GSPDB:GN00116; NCSP:B7F21.50

A:Experimental source: BAC clone B7F21; strain OR74A

C:Genetics:

A:Gene: NCSP:B7F21.50

A:Map position: 6

A:Introns: 117/1

Query Match 12.4%; Score 96; DB 2; Length 770;
Best Local Similarity 22.8%; Pred. No. 1.8;
Matches 31; Conservative 19; Mismatches 54; Indels 32; Gaps 3;

QY 28 GCGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADN 87
DB 389 GAGNNNGSELASDPNRYFD---GASPPIPSSSHSTTSTTYAGGHNGHGMRESVDS 444

QY 88 STIEL-----TQNGFRNNATIDQWNAKNSDITVGQYDQLV----- 122
DB 445 QSTGLGVHYSSTTRNQHHWRNQSNASELSADGSEITHGVASPLVGGSSHARGASGGTS 504

QY 123 ---TRVVTHEMAHASV 135

DB 505 YRYTHTHSHSHSGL 520

RESULT 12

S42136

cnjB protein - Tetrahymena thermophila

C:Species: Tetrahymena thermophila

C>Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999

C:Accession: S42136; S03650

R:Taylor, P.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A:Reference number: S42136

A:Accession: S42136

A:Molecule type: DNA
 A:Residues: 1-1748 <RAY>
 A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752
 R:Taylor, F.M.; Martindale, D.W.
 Nucleic Acids Res. 21, 4610-4614, 1993
 A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
 A:Reference number: S42135; MUID:94051569; PMID:8233798
 A:Accession: S42135
 A:Molecule type: DNA
 A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1
 A:Cross-references: EMBL:L03710
 R:Taylor, F.M.; Martindale, D.W.; Taylor, F.M.
 Nucleic Acids Res. 16, 2189-2201, 1988
 A:Title: Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*.
 A:Reference number: S03650; MUID:88189811; PMID:3357771
 A:Accession: S03650
 A:Molecule type: DNA
 A:Residues: 236-250, 'I', 252-255, 'N', 257-773 <MAR>
 A:Cross-references: EMBL:X06462
 C:Genetics:
 A:Gene: cniB
 A:Genetic code: SGC5
 A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 708/3; 777/3; 8
 C:Keywords: zinc finger
 F:1164-1450/Region: glycine-rich
 F:1451-1464/Region: zinc finger CCHC motif
 F:1478-1491/Region: zinc finger CCHC motif
 F:1501-1514/Region: zinc finger CCHC motif
 F:1530-1543/Region: zinc finger CCHC motif
 F:1555-1568/Region: zinc finger CCHC motif
 F:1579-1592/Region: zinc finger CCHC motif
 F:1602-1615/Region: zinc finger CCHC motif
 F:1626-1748/Region: glycine-rich

Query Match 12.3%; Score 95; DB 2; Length 1748;
 Best Local Similarity 32.0%; Pred. No. 5.6;
 Matches 31; Conservative 11; Mismatches 29; Indels 26; Gaps 4;
 A:Map position: 2

QY 25 QWGGGNGHNGA---GNSGSPDSTLSIYQGSANAALQSDARKSETIT---QSGYGN 77
 DB 1640 QFGGGNSNGQSGNGTSSGSDWN-----CQSNVQESTTSSGGWSSGSGN 1685

QY 78 GADVGGQAGDNSTIELTQNGFRNNATIDQNAKNSDIT 114
 DB 1686 QTGGGWSNDN-----CQQQNTGGGGWSNSNQ 1717

RESULT 13
 E5965
 Hypothetical glycine-rich protein [imported] - *Sinorhizobium meliloti* (strain 1021) mega
 C:Species: *Sinorhizobium meliloti*
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E5965
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-Rb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E5965
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2174 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CA49389.1; PID:g15140875; GSPDB:GN00167
 R:Experimental source: strain 1021, megaplasmid pSymB
 R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
 pla, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
 heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21369234; PMID:11474104
 A:Contents: annotation
 C:Genetics:

A:Gene: SMD21548
 A:Genome: plasmid

Query Match 12.0%; Score 93; DB 2; Length 2174;
 Best Local Similarity 24.5%; Pred. No. 11;
 Matches 36; Conservative 15; Mismatches 32; Indels 64; Gaps 7;
 A:Map position: 2

QY 11 AIVVSGSALAGVVPQ--WGGGNGHNGGNSGSPDSTLSIYQGSANAALQSDARKSET 68
 DB 693 AIATAGAGAVGILASIGGGGN---GGNATGGAGFGSFGGGGGG----- 737
 QY 69 TITQSGYNGADVG--QGADNSTIELTQNGFRNNATIDQNAKNSDITVQYDQLVTRVVT 127
 DB 738 ---GGYANTANVFGK-----LTLTQG----- 757
 QY 128 HEMAHASVMVRQV---GFGNATANQY 151
 DB 758 ---SHAAAGIVAQSVGGGGTGGTASSY 781

RESULT 14
 B84533
 Hypothetical protein At2g15770 [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cross)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: B84533
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: B84533
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-301 <STO>
 A:Cross-references: GB:AE002093; NID:G5306254; PIDN:AA41987.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g15770
 A:Map position: 2

Query Match 11.7%; Score 91; DB 2; Length 301;
 Best Local Similarity 26.4%; Pred. No. 1.7;
 Matches 33; Conservative 19; Mismatches 51; Indels 22; Gaps 5;
 A:Map position: 2

QY 15 SGSALA-----GVVPQWGGGNGH--NGGNSGSPDSTLSIYQGSANAALQSDARKSET 68
 DB 69 SGGGLGNSNNGSGWGWGTGSGSGSGSTNPDCGRRSRWNW-----SLKSGWSWG 121

QY 69 TITQSGYNGADVGQAGDNST---IELTQNGFRNNATIDQNAKNSDITVQYD 119
 DB 122 SNDNDSNNGSGSDSGGLDRETPKTIIVGGSDGWKGLDYKDWSKNAPFYVNDLVFKYD 181

QY 120 OLVT 124
 DB 182 KSAKR 186

RESULT 15
 A56038
 DNA-binding protein ovo - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
 C:Accession: A56038
 R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
 Mol. Cell. Biol. 14, 6809-6818, 1994
 A:Title: Multiple products from the shavenbaby-ovo gene region of *Drosophila melanogast*
 A:Reference number: A56038; MUID:95021209; PMID:7935398
 A:Accession: A56038
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1028 <GAR>
 A:Cross-references: GB:U11383; NID:G520526; PIDN:AAB60216.1; PID:G520527

C;Genetics:
A;Gene: ovo
A;Cross-references: FlyBase:FBgn0003028

Query Match 11.7%; Score 91; DB 2; Length 1028;
Best Local Similarity 31.2%; Pred. No. 6.7;
Matches 34; Conservative 8; Mismatches 41; Indels 26; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVFPQWGGGNNHGGNSGPDSTLSIYQYGSANAALQSD 62
Db 59 LQNAAAAAYIMSAGSG-----GGCTGNGGGGASGPGGSPSANSGGGGGG----- 104

QY 63 ARKSETTITQSGYNGADYQGADNSTIELTONGFRNNATIDOWNAKNS 111
Db 105 -----CGNGYINCGVG-GPNNs---LDGNLLNFASVSNYNESNS 141

Search completed: August 2, 2004, 14:56:21
Job time : 9.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 693 | 89.4 | 151 | 1 | CSGA_SALTY |
| 2 | 528 | 68.1 | 151 | 1 | CSGA_ECOLI |
| 3 | 506.5 | 65.4 | 152 | 1 | CSGA_BCO57 |
| 4 | 104.5 | 13.5 | 151 | 1 | CSGB_ECOLI |
| 5 | 98.5 | 12.7 | 151 | 1 | CSGB_SALTY |
| 6 | 92 | 11.9 | 1327 | 1 | TNKL_HUMAN |
| 7 | 91 | 11.7 | 1028 | 1 | OVO_DROME |
| 8 | 90.5 | 11.7 | 151 | 1 | CSGB_SALTY |
| 9 | 90 | 11.6 | 1093 | 1 | PER_DROWI |
| 10 | 89.5 | 11.5 | 262 | 1 | VG38_BPT2 |
| 11 | 88.5 | 11.4 | 479 | 1 | PRTC_ERWCH |
| 12 | 87 | 11.2 | 646 | 1 | GP63_LEIME |
| 13 | 87 | 11.2 | 678 | 1 | YF48_MYCTU |
| 14 | 86 | 11.1 | 599 | 1 | GP63_LEICH |
| 15 | 85.5 | 11.0 | 1048 | 1 | AG01_ARATH |
| 16 | 85 | 11.0 | 347 | 1 | MSA2_PLAF2 |
| 17 | 84 | 10.8 | 590 | 1 | GP63_LEIDO |
| 18 | 84 | 10.8 | 602 | 1 | GP63_LEIMA |
| 19 | 83.5 | 10.8 | 1460 | 1 | PD46_CABEL |
| 20 | 83.5 | 10.8 | 1067 | 1 | SGG_DROME |
| 21 | 83 | 10.7 | 362 | 1 | P35_MYCPE |
| 22 | 83 | 10.7 | 1656 | 1 | OMP8_RICJA |
| 23 | 82 | 10.6 | 1080 | 1 | HDC_DROME |
| 24 | 81.5 | 10.5 | 392 | 1 | HMET_HUMAN |
| 25 | 81 | 10.5 | 165 | 1 | GR1_ORYSA |
| 26 | 81 | 10.5 | 1612 | 1 | RRPO_PMWUJ |
| 27 | 81 | 10.5 | 1612 | 1 | RRPO_PMWVS |
| 28 | 80.5 | 10.4 | 481 | 1 | PRTB_ERWCH |
| 29 | 80.5 | 10.4 | 641 | 1 | IMD_ARTGO |
| 30 | 80 | 10.3 | 385 | 1 | PER_DROME |
| 31 | 80 | 10.3 | 491 | 1 | YK38_MYCTU |
| 32 | 80 | 10.3 | 548 | 1 | CEAK_ECOLI |
| 33 | 79.5 | 10.3 | 204 | 1 | HEVE_HEVER |

ALIGNMENTS

RESULT 1

| ID | CSGA_SALTY | STANDARD | PRT | 151 AA |
|------|--|-----------------------------------|-----|--------|
| AC | P52257 | | | |
| DT | 01-OCT-1996 | (Rel. 34, Created) | | |
| DT | 01-OCT-1996 | (Rel. 34, Last sequence update) | | |
| DT | 10-OCT-2003 | (Rel. 42, Last annotation update) | | |
| DE | Major curlin subunit precursor (Fimbrin SEF17). | | | |
| CSGA | OR AGPA OR STM144 OR STY1181 OR T1776. | | | |
| GN | Salmonella typhimurium, | | | |
| OS | Salmonella typhi, and | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | | |
| OC | Enterobacteriaceae; Salmonella. | | | |
| CX | NCBI_TaxID=602, 601, 592; | | | |
| [1] | SEQUENCE FROM N.A. | | | |
| RP | SPECIES=S.typhimurium; STRAIN=SR-11; | | | |
| RC | MEDLINE=98117058; PubMed=9457880; | | | |
| RX | Rowling U., Bian Z., Hammar M., Sierralta W.D., Normark S.; | | | |
| RA | "Curli fibers are highly conserved between Salmonella typhimurium and | | | |
| RT | Escherichia coli with respect to operon structure and regulation."; | | | |
| RL | J. Bacteriol. 180:722-731(1998). | | | |
| [2] | SEQUENCE FROM N.A. | | | |
| RP | SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720; | | | |
| RC | MEDLINE=21534948; PubMed=11677609; | | | |
| RX | McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., | | | |
| RA | Courneý L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., | | | |
| RA | Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., | | | |
| RA | Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., | | | |
| RA | Waterston R., Wilson R.K.; | | | |
| RT | "Complete genome sequence of Salmonella enterica serovar Typhimurium | | | |
| RL | LT2."; | | | |
| RL | Nature 413:852-856(2001). | | | |
| [3] | SEQUENCE FROM N.A. | | | |
| RP | SPECIES=S.typhi; STRAIN=CT18; | | | |
| RC | MEDLINE=21534947; PubMed=11677608; | | | |
| RX | Parkhill J., Dougan G., James K.D., Holden M.T.G., Sebatina M., | | | |
| RA | Churcher C., Mungall K.L., Bentley K., Chillingworth T., Connor P., | | | |
| RA | Baker S., Basham D., Brooks K., Davies R.M., Dowd L., White N., Farrar J., | | | |
| RA | Cronin A., Davis P., Davies R.M., Hien T.T., Holroyd S., Jagels K., | | | |
| RA | Feltwell T., Hamlin N., Haque A., Kien T.T., Holroyd S., Jagels K., | | | |
| RA | Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Barry C., | | | |
| RA | Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., | | | |
| RA | Whitehead S., Barrall B.G.; | | | |
| RT | "Complete genome sequence of a multiple drug resistant salmonella | | | |
| RL | enterica serovar Typhi CT18."; | | | |
| RL | Nature 413:848-852(2001). | | | |
| [4] | SEQUENCE FROM N.A. | | | |
| RP | SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931; | | | |
| RC | MEDLINE=22531367; PubMed=12644504; | | | |
| RX | Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., | | | |
| RA | Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.; | | | |

P91698 drosophila
Q8xf0 homo sapien
P18127 xanthomonas
P30688 neisseria m
Q24180 drosophila
Q04893 saccharomyc
P34291 caenorhabdi
P34308 caenorhabdi
P52302 drosophila
Q12756 homo sapien
P06997 escherichia
P55004 parbitis n

34 79 10.2 396 1 PER_DROPV
35 79 10.2 933 1 NPA3_HUMAN
36 78.5 10.1 1567 1 ICEN_XANCT
37 78 10.1 331 1 OMB2_NEIMB
38 78 10.1 576 1 DEAF_DROME
39 77.5 10.0 1140 1 YM96_YEAST
40 77 9.9 401 1 YK03_CABEL
41 77 9.9 783 1 YK02_CABEL
42 77 9.9 796 1 MBN_DROME
43 77 9.9 1690 1 KFLA_HUMAN
44 76.5 9.9 329 1 SUBI_ECOLI
45 76.5 9.9 1115 1 PHYE_PHANI

"Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";
 J. Bacteriol. 185:2330-2337(2003).
 [5] SEQUENCE FROM N.A.
 RP SPECIES=enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6] SEQUENCE OF 21-151 FROM N.A.
 RP SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
 RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
 RT the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7] SEQUENCE OF 21-33.
 RP SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 RN CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ002301; CA05317.1; -;
 DR EMBL; AE008749; RAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; AE016840; AAO69399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR PIR; JC6039; JC6039.
 DR StyGene; SG10608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMVQVGFGNATANYQ -> DSVTQVAS (IN
 FT REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 89.4%; Score 693; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.9e-53;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVSGSALAGVVPWGGGNGHNGGSGPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVSGSALAGVVPWGGGNGHNGGSGPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITSGYNGADVGQADNSTIETQNGFRNATIDQWAKNSDITVGYDQ 120
 DB 61 SDARKSETTITSGYNGADVGQADNSTIETQNGFRNATIDQWAKNSDITVGYDQ 120
 QY 121 LVTRVVTHEMAHASVWVQVGFGNATANYQ 151
 DB 121 NNAALVNQATSDSSVWVQVGFGNATANYQ 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 RT repression of csGA, the subunit gene of fibronectin-binding curli in
 RT Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2] SEQUENCE FROM N.A.
 RP STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli
 RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3] SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4] SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5] SEQUENCE OF 21-40.
 RP STRAIN=K12 / YMEL;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 RT fibronectin binding in Escherichia coli HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6] SEQUENCE OF 21-31.
 RP MEDLINE=91310586; PubMed=1677357;
 RX Collinson S.K., Emedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RT Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.

```

CC  -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; L04979; AAA23616.1; -.
DR  EMBL; X90754; AAC62282.1; -.
DR  EMBL; AF000205; AAC74126.1; -.
DR  EMBL; D90741; BAA35832.1; -.
DR  EMBL; D90742; BAA35840.1; -.
DR  PIR; S70788; S70788.
DR  EcoGene; EGI1489; csGA.
KW  Fimbria; Signal; Complete proteome.
FT  SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
FT CHAIN 1 20
FT CONFLICT 7 7 A -> E (IN REF. 1). CRC64;
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 68.1%; Score 528; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 7,9e-39;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAATAAIVFSSALAGVVPQYGGGNGHGGNSGPNSEINLYYQGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGVQYDQ 120
DB 61 TDARNSDLTITQHGGNGADVGQSDSDSIDLTQRFGNSATLDQWNGKNSMTVKQFGG 120

QY 121 LVTRVVTHEMAHASVMVQVGFNNATANQY 151
DB 121 GNAAVVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID CSGA_ECO57
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]_TaxID=83334;
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elger R.O.;
RA "Mutations in the csGD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RT Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / BDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Foster G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoulsis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).

```

```

RN [3] SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Havaashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; AF275733; AAK53212.1; -.
DR  EMBL; AE005315; AAG55788.1; -.
DR  EMBL; AP002554; BAB34843.1; -.
DR  PIR; D90806; D90806.
DR  PIR; H85665; H85665.
KW  Fimbria; Signal; Complete proteome.
FT  SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 1 20 MAJOR CURLIN SUBUNIT.
FT CONFLICT 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 65.4%; Score 506.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 5,6e-37;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQGSANAALAL 59
DB 1 MKLLKVAATAAIVFSSALAGVVPQYGGGNGHGGNSGPNSEINLYYQGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGVQYD 119
DB 61 QADARNSDLTITQHGGNGADVGQSDSDSIDLTQRFGNSATLDQWNGKNSMTVKQFG 120
QY 120 QLVTRVVTHEMAHASVMVQVGFNNATANQY 151
DB 121 GNAAVVDQTASNSVNVTVQVGFNNATAHQY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]_TaxID=562, 83334;
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in Escherichia coli

```

```

DR ENBL; AE000205; AAC74125.1; -.
DR ENBL; D90741; BAA35831.1; -.
DR ENBL; AE005315; AAG55787.1; -.
DR ENBL; AP002554; BAB34842.1; -.
DR PIR; C90806; C90806.
DR PIR; G85655; G85655.
DR PIR; S70787; S70787.
DR EcoGene; EG12621; csbB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match 13.5%; Score 104.5; DB 1; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.023;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FFAIVVGSAL--AGVVPQWGGGNGHNGSGSPDSTLSIY-QYGSANAALALOSDARK 65
DB 33 FAYNELSKSFNCAII--GQAGTNNSAQLQGGSKLLAVQAEGSSNRA-KIDQTGDY 88

QY 66 SETTITQSGYGVGADVQCGADNSITIELTONGFPFNNTATQWNAKNSDITVGVYDQLVTRV 125
DB 89 NLAYIDQAGSANDASISQAYGNTAMIIQKSGNKANITQYGTOKTAIVVQRQSQWAIEV 148

RESULT 5
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY STANDARD; AC
P52226; AC
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
CSGB OR AGFE OR STM143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=602, 592;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli" fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RT J. Bacteriol. 180:722-731(1998).
RN [2]
SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea B., Miller W., Stoneking I., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RN [3]
Nature 413:852-856(2001).
RN [3]
SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8850497;
RA Collinson S.K., Clouthier S.C., Doran J.B., Bauser P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RN [3]
J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.

```



```

DR EMBL; U51065; AAB41370.1; -
DR EMBL; U51066; AAB41371.1; -
DR EMBL; U51067; AAB41372.1; -
DR EMBL; U51068; AAB41373.1; -
DR EMBL; U51069; AAB41374.1; -
DR EMBL; U51070; AAB41375.1; -
DR EMBL; U51071; AAB41376.1; -
DR EMBL; U51072; AAB41377.1; -
DR EMBL; L06342; AAA28765.1; -
DR FlyBase; FBgn0013161; Dm11per.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS0112; PAS; 2.
KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation;
KW Polymorphism.
FT NON_TER 1
FT DOMAIN <1 12 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 139 209 PAS 1.
FT DOMAIN 289 359 PAS 2.
FT DOMAIN 371 411 PAC.
FT DOMAIN 7 12 POLY-LYS.
FT DOMAIN 618 625 POLY-GLY.
FT DOMAIN 718 734 POLY-GLY.
FT DOMAIN 745 748 POLY-SER.
FT DOMAIN 759 770 POLY-GLY.
FT DOMAIN 885 888 POLY-ALA.
FT DOMAIN 911 917 POLY-ALA.
FT VARIANT 611 611 T -> A (IN STRAIN 0811.4).
FT VARIANT 617 617 S -> F (IN STRAIN 0811.4).
FT VARIANT 622 622 G -> V (IN STRAIN GUANA).
FT VARIANT 724 724 G -> A (IN STRAIN MANAUS 2).
FT VARIANT 726 726 G -> S (IN STRAIN SANTA MARIA).
FT VARIANT 729 734 MISSING (IN STRAIN PORTO ALEGRE 3).
FT VARIANT 730 734 MISSING (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 731 734 MISSING (IN STRAINS MANAUS 1 AND MANAUS 3).
FT VARIANT 732 734 MISSING (IN STRAINS LIMA B, L'HABITATUE AND CANO MORA).
FT VARIANT 733 733 G -> V (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 733 734 MISSING (IN STRAINS GUADELOUPE, MANAUS 2, PORTO ALEGRE 2, PORTO ALEGRE 1 AND GUANA).
FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO ALEGRE 1 AND PORTO ALEGRE 2).
FT VARIANT 747 747 S -> A (IN STRAINS GUADELOUPE AND GUANA).
FT VARIANT 764 766 MISSING (IN STRAIN MANAUS 3).
FT VARIANT 886 886 A -> T (IN STRAIN 0811.4).
FT NON_TER 1093 1093
SQ SEQUENCE 1093 AA; 115896 MW; AB6DR050267EC187 CRC64;

Query Match 11.6%; Score 90; DB 1; Length 1093;
Best Local Similarity 27.7%; Pred. No. 3.9; Indels 16; Gaps 2;
Matches 26; Conservative 6; Mismatches 46;

QY 27 GGGGHHNGGSSGDPSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGAD 86
DB 721 GGGGGGGGGGGGLFLFLDVTHTSS-----SSQNKGPVGAAGGAGGGVGGGG-- 770

QY 87 NSTIELTQNGFRNATIDOWNAKNSDITVGYDQ 120
DB 771 -----SCSLGNGNGVSGNGNNSQFSTNQYQ 798

RESULT 10
VG38_BPT2 STANDARD; PRT; 262 AA.
AC P07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.

OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OC NCBI_TaxID=10664;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=87283911; PubMed=3302276;
RX Riede I., Drexler K., Eschbach M.L., Henning U.;
RA "DNA sequence of genes 38 encoding a receptor-recognizing protein of
RT bacteriophages T2, K3 and of K3 host range mutants.";
RL J. Mol. Biol. 194:31-39(1987).
CC -!- FUNCTION: VG38 is at the tip of the long tail fibers and serves as
CC the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
CC AS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05312; CAA28935.1; -
DR PIR; S00275; S00275; Tail_fibre_GP38.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
DR KW Fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 262;
Best Local Similarity 34.1%; Pred. No. 0.83; Indels 17; Gaps 4;
Matches 30; Conservative 8; Mismatches 33;

QY 27 GGGGHHNGGSSGDPSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGAD 86
DB 175 GGGGRPFVGKGIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGNGVGGGG 225

QY 87 NSTIELTQNGFRNATIDOWNAKNSDIT 114
DB 226 -----RGWKNVYTSFGAAGAAVT 245

RESULT 11
PRTC_ERWCH STANDARD; PRT; 479 AA.
AC P16317;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Secreted protease C precursor (EC 3.4.24.-) (Proc).
GN PRTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=B374;
RX MEDLINE=91009140; PubMed=2211614;
RA Delepeleire P., Wandersman C.;
RT "Protein secretion in Gram-negative bacteria. The extracellular
RT metalloprotease B from Erwinia chrysanthemi contains a C-terminal
RT secretion signal analogous to that of Escherichia coli alpha-
RT hemolysin."
RL J. Biol. Chem. 265:17118-17125(1990).
RN [2]
RP SEQUENCE OF 1-59 FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=B374;
RX MEDLINE=89255387; PubMed=2722818;
RA Delepeleire P., Wandersman C.;
RT "Protease secretion by Erwinia chrysanthemi. Proteases B and C are

```

FT synthesized and secreted as zymogens without a signal peptide.";
 RL J. Biol. Chem. 264:9083-9089(1989).
 CC -|- COPACTOR: Binds 1 zinc ion and 7 calcium ions per subunit (By
 CC similarity).
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- DOMAIN: THE GLY-RICH REPEATS MAY BE IMPORTANT IN THE EXTRACELLULAR
 CC SECRETION OF THIS METALLOPROTEASE.
 CC -|- SIMILARITY: Belongs to peptidase family M10B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M59229; AAA24860.1; -;
 DR EMBL; J04736; AAA24862.1; -;
 DR EMBL; M60395; AAA63638.1; -;
 DR PIR; A38307; A38307.
 DR PDB; 1GO7; 17-OCT-02.
 DR PDB; 1GO8; 17-OCT-02.
 DR PDB; 1K7G; 19-OCT-02.
 DR PDB; 1K7Q; 20-OCT-02.
 DR MEROPS; M10.054; -;
 DR InterPro; IPR001343; Hemolysin_Ca_bind.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR006026; Peptidase_M.
 DR Pfam; PF00353; hemolysinCbind; 3.
 DR PRINTS; PR00313; CASNDNGRPT.
 DR SMART; SMO0235; ZnMc; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00330; HEMOLYSIN_CALCULUM; 1.
 KW Hydrolyase; Metalloprotease; Calcium-binding; Metal-binding; Zinc;
 KW Repeat; Zymogen; 3D-structure.
 FT PROPEP 1 17
 FT CHAIN 18 479 SECRETED PROTEASE C.
 FT REPEAT 345 353 GGXGD 1.
 FT REPEAT 354 362 GGXGD 2.
 FT REPEAT 363 371 GGXGD 3.
 FT REPEAT 372 380 GGXGD 4.
 FT REPEAT 381 389 GGXGD 5.
 FT METAL 188 188 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 189 189 BY SIMILARITY.
 FT METAL 192 192 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 265 265 CALCULUM 1 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 267 267 CALCULUM 1 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 297 297 CALCULUM 1 (BY SIMILARITY).
 FT METAL 299 299 CALCULUM 1 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 300 300 CALCULUM 2 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 302 302 CALCULUM 1 AND 2 (BY SIMILARITY).
 FT METAL 339 339 CALCULUM 2 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 341 341 CALCULUM 2 (BY SIMILARITY).
 FT METAL 346 346 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 348 348 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 350 350 CALCULUM 3 (BY SIMILARITY).
 FT METAL 355 355 CALCULUM 4 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 357 357 CALCULUM 4 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 359 359 CALCULUM 4 (BY SIMILARITY).
 FT METAL 363 363 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 364 364 CALCULUM 5 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).

FT METAL 365 365 SIMILARITY).
 FT METAL 366 366 CALCULUM 3 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 368 368 CALCULUM 3 AND 5 (BY SIMILARITY).
 FT METAL 372 372 CALCULUM 4 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 373 373 CALCULUM 6 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 375 375 CALCULUM 6 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 377 377 CALCULUM 4 AND 6 (BY SIMILARITY).
 FT METAL 381 381 CALCULUM 5 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 382 382 CALCULUM 7 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 383 383 CALCULUM 5 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 384 384 CALCULUM 7 (VIA CARBONYL OXYGEN) (BY
 FT SIMILARITY).
 FT METAL 386 386 CALCULUM 5 AND 7 (BY SIMILARITY).
 FT METAL 395 395 CALCULUM 6 (BY SIMILARITY).
 FT METAL 402 402 CALCULUM 6 (BY SIMILARITY).
 FT METAL 412 412 CALCULUM 7 (BY SIMILARITY).
 SQ SEQUENCE 479 AA; 51600 MW; 990ED8376725DF61 CRC64;
 Query Match 11.4%; Score 88.5; DB 1; Length 479;
 Best Local Similarity 28.8%; Pred. No. 2;
 Matches 40; Conservative 21; Mismatches 41; Indels 37; Gaps 9;
 QY 27 GGGHHNGGSGSDSTLSIYQGSANNAALQDARKSETITQSG-----YNG----- 78
 DB 345 GSGNDILVNSAD-----NILQGGAGNDLVYGGAGA---DTLYGGAGRDITVYSGGQS 396
 QY 79 -----ADVGGAGDNSTIETONGFRNNATII---DOWNAKNSDITVGYD--CLVTR 124
 DB 397 TVAYDWIADFOKGDID---KIDLSAFRNEGQLSFVQDFTGKGQEVML-QWDANSIIN 451
 QY 125 VVTHEVAHASV--MVRQVG 141
 DB 452 LMLHEAGHSVDLVRIVG 470
 RESULT 12
 GP63_LEIME ID GP63_LEIME STANDARD; PRT; 646 AA.
 AC P43150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 OS GP63-C1.
 OS Leishmania mexicana.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MNYC/BZ/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Kares R.E., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania
 RT mexicana are developmentally regulated";
 RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -|- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -|- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
 CC -|- COPACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -|- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the


```

RT "Empirical analysis of transcriptional activity in the Arabidopsis
RL genome.";
RL Science 302:842-846(2003).
CC -!- FUNCTION: Essential for proper development of leaves and floral
CC organs, and formation of axillary meristems.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. According to EST
CC sequences;
CC Name=1;
CC IsoId=O04379-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Widely expressed at low levels.
CC -!- DEVELOPMENTAL STAGE: Expressed throughout all developmental
CC stages.
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC -!- CAUTION: Ref.2 (AAF79718) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U91995; AAC18440.1; -
CC EMBL; AC007932; AAD49755.1; -
CC EMBL; ACC20889; AAF79718.1; ALT_SEQ.
CC EMBL; BT000941; AAN41341.1; -
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS50821; PAZ; 1.
CC PROSITE; PS50822; PIWI; 1.
CC Developmental protein; Alternative splicing.
CC FT DOMAIN 391 501 PAZ.
CC FT DOMAIN 676 997 PAZ.
CC FT DOMAIN 13 104 GLY-RICH.
CC SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;
Query Match 11.0%; Score 85.5; DB 1; Length 1048;
Best Local Similarity 25.7%; Pred.No.9.2;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;
QY 27 GGGGNHNGGNSGGPDSTLSI-----YQGSANAALALQSDARKSETTITQSGYG 76
Db 95 GYGGRGGGPGSPGPPQRCSPVELHQATSPYQAVSSQPTLSEVSPQTVPETVLAQGF- 153
QY 77 NGADYVQQGADNSTIE-----LTQNGFERNATTIDOWNAKNSDITVQGYDQ 120
Db 154 EQLSV EQGAPSOAIQPIFSSSKAFKFPMPRPGKSGKRCIVKANHFFAELEPKDLHHYDV 213
QY 121 LVTRVVTHEMAHASVMVRCV 140
Db 214 TITPEVTSRGVNRVANKQLV 233

```

Search completed: August 2, 2004, 14:49:28
Job time : 6.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)

1604.150 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*
1: sp_arChaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mbc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|--------|--------------------|
| 1 | 684 | 88.3 | 152 | Q33802 | Q33802 salmonella |
| 2 | 578.5 | 74.6 | 150 | Q7X243 | Q7X243 citrobacter |
| 3 | 533 | 68.8 | 149 | Q7X240 | Q7X240 citrobacter |
| 4 | 503.5 | 65.0 | 152 | Q8CW63 | Q8CW63 escherichia |
| 5 | 438.5 | 56.6 | 150 | Q7X237 | Q7X237 enterobacte |
| 6 | 385 | 49.7 | 76 | Q54069 | Q54069 salmonella |
| 7 | 122 | 15.7 | 29 | Q9S3J5 | Q9S3J5 escherichia |
| 8 | 109 | 14.1 | 179 | Q33801 | Q33801 salmonella |
| 9 | 107.5 | 13.9 | 151 | Q7X238 | Q7X238 enterobacte |
| 10 | 105 | 13.5 | 139 | Q8EIH3 | Q8EIH3 shewanella |
| 11 | 104.5 | 13.5 | 152 | Q7X241 | Q7X241 citrobacter |
| 12 | 104.5 | 13.5 | 160 | Q8CW64 | Q8CW64 escherichia |
| 13 | 104 | 13.4 | 502 | Q8EIH4 | Q8EIH4 shewanella |
| 14 | 103.5 | 13.4 | 151 | Q7X244 | Q7X244 citrobacter |
| 15 | 103.5 | 13.4 | 151 | Q7UC21 | Q7UC21 shigella fl |
| 16 | 103.5 | 13.4 | 160 | Q83RU7 | Q83RU7 shigella fl |

| | | | | | | |
|----|------|------|------|----|--------|--------------------|
| 17 | 101 | 13.0 | 362 | 16 | Q8EV84 | Q8EV84 mycoplasma |
| 18 | 99.5 | 12.8 | 348 | 13 | Q93397 | Q93397 cyprinus ca |
| 19 | 96.5 | 12.5 | 552 | 16 | P96840 | P96840 mycobacteri |
| 20 | 96.5 | 12.5 | 552 | 16 | Q7W76 | Q7W76 mycobacteri |
| 21 | 96.5 | 12.5 | 523 | 16 | Q8VIY0 | Q8VIY0 mycobacteri |
| 22 | 96.5 | 12.5 | 1422 | 16 | Q8EFU3 | Q8EFU3 shewanella |
| 23 | 96.5 | 12.5 | 3501 | 16 | Q8YI06 | Q8YI06 raistonia s |
| 24 | 96.5 | 12.5 | 3552 | 16 | Q8XSD6 | Q8XSD6 raistonia s |
| 25 | 96 | 12.4 | 154 | 16 | Q89JIS | Q89JIS bradyrhizob |
| 26 | 96 | 12.4 | 157 | 16 | Q8EHG0 | Q8EHG0 pseudomonas |
| 27 | 96 | 12.4 | 606 | 3 | Q9P3I9 | Q9P3I9 neurospora |
| 28 | 95.5 | 12.3 | 824 | 3 | Q8NIV1 | Q8NIV1 neurospora |
| 29 | 95.5 | 12.3 | 1785 | 16 | Q7V8S5 | Q7V8S5 prochloroco |
| 30 | 95 | 12.3 | 1748 | 5 | Q94821 | Q94821 tetrahymena |
| 31 | 94.5 | 12.2 | 191 | 3 | Q8TFA6 | Q8TFA6 saccharomyc |
| 32 | 94 | 12.1 | 171 | 16 | Q89JI3 | Q89JI3 bradyrhizob |
| 33 | 94 | 12.1 | 368 | 16 | Q8EWD6 | Q8EWD6 mycoplasma |
| 34 | 93 | 12.0 | 480 | 16 | Q89EV2 | Q89EV2 bradyrhizob |
| 35 | 93 | 12.0 | 2174 | 16 | Q92UU8 | Q92UU8 rhizobium m |
| 36 | 92 | 11.9 | 453 | 5 | Q9N6M8 | Q9N6M8 drosophila |
| 37 | 91 | 11.7 | 301 | 10 | Q9XIL0 | Q9XIL0 arabidopsis |
| 38 | 91 | 11.7 | 453 | 5 | Q9NGF6 | Q9NGF6 drosophila |
| 39 | 91 | 11.7 | 453 | 5 | Q9NGF7 | Q9NGF7 drosophila |
| 40 | 91 | 11.7 | 1222 | 5 | Q9W4F0 | Q9W4F0 drosophila |
| 41 | 91 | 11.7 | 1222 | 5 | Q8T8L9 | Q8T8L9 drosophila |
| 42 | 91 | 11.7 | 1551 | 5 | Q8SX56 | Q8SX56 drosophila |
| 43 | 91 | 11.7 | 1354 | 5 | Q8MPN4 | Q8MPN4 drosophila |
| 44 | 90.5 | 11.7 | 145 | 16 | Q8U6N9 | Q8U6N9 agrobacteri |
| 45 | 90.5 | 11.7 | 191 | 10 | Q7XDR3 | Q7XDR3 oryza sativ |

ALIGNMENTS

RESULT 1

O33802 : PRELIMINARY; PRT; 152 AA.
ID O33802 :
AC O33802 :
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AgfA protein (Fragment).
AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CAA04151.1; -;
FT NON TER 152
FT SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;
SQ

Query Match 88.3%; Score 684; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 1.2e-47;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

| | | | | | | | |
|----|-----|------------------------|------------|-------------|----------|-------------|----------|
| QY | 1 | MKLLKVAFAAIVVSGSALA | VVVG | GGGNGGNGSSG | SPDSTLSI | YVGSANAALAQ | 60 |
| Db | 1 | MKLLKVAFAAIVVSGSALA | VVVG | GGGNGGNGSSG | SPDSTLSI | YVGSANAALAQ | 60 |
| QY | 61 | SDARKSETTITGSGYNGADVGQ | GDADN | STIELTQNGFR | NNATIDQ | WNAKNSDITV | QYDQ 120 |
| Db | 61 | SDARKSETTITGSGYNGADVGQ | GDADN | STIELTQNGFR | NNATIDQ | WNAKNSDITV | QYDQ 120 |
| QY | 121 | LIVTRVTHEMAHASMVRQV | FGNNATANQY | 151 | | | |


```

OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rec39;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515702; CAD56677.1; -.
SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A27882B7 CRC64;

Query Match 13.9%; Score 107.5; DB 2; Length 151;
Best Local Similarity 30.8%; Pred. No. 0.31;
Matches 36; Conservative 14; Mismatches 42; Indels 25; Gaps 4;

QY 10 AATVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSETT 69
DB 58 AQIQBGSKLLSVSQ-----DGAGNEARVD-----QSGTYNIAMIDQS-GNGNDAG 103

QY 70 ITQSGYNGADVGQAGNADNSTIELTQNGFNFNATIDQWNAKNSDITVGYDQLVTRV 126
DB 104 ITQDGYGNSA-----XIIQKSGNRRNITQYGTQKTAVVVQSQQWAIHV 149

RESULT 10
Q8EIH3 PRELIMINARY; PRT; 139 AA.
ID Q8EIH3
AC Q8EIH3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN SC08666.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2257686; PubMed=12368913;
RA Heidelberg J.F., Paulsen I.T., Neilson K.B., Gaidos E.J., Neilson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Mathe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; A2015532; AAN53942.1; -.
DR TIGR; SC08666; -.
KW Complete Proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 13.5%; Score 105; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.45;
Matches 32; Conservative 23; Mismatches 44; Indels 14; Gaps 4;

QY 39 SGPDSLTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGQAGNADNSTIELTQNGFR 98
DB 41 SGRDNIIDLVQQTANQGVIVFGSGDNS--AYVTQAGNDNISLVTQIGTNNVQLLQVGAQ 99

QY 99 NNATIDQWNAKNSDITVGYDQLVTRVTHMAHASVMVRQVGFNNATANQY 151
DB 100 NKASITQ-----IGN-DNLVQ---LNQLGSGNFSIQIADGAASITQY 139

RESULT 11
Q7X241 PRELIMINARY; PRT; 152 AA.
ID Q7X241
AC Q7X241;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pec4;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56674.1; -.
SQ SEQUENCE 152 AA; 16149 MW; D063A527D45D4329 CRC64;

Query Match 13.5%; Score 104.5; DB 2; Length 152;
Best Local Similarity 29.1%; Pred. No. 0.55;
Matches 32; Conservative 13; Mismatches 40; Indels 25; Gaps 3;

QY 16 GSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGY 75
DB 65 GSXLLSVISQ-----EGSGNEAKTDQT-----GSGNFAY-----IDQTGS 99

QY 76 GNGADVGQAGNADNSTIELTQNGFNFNATIDQWNAKNSDITVGYDQLVTRV 125
DB 100 SNDASIKQSGYNGTAVIIQKSGNKNANITQYGTQKTAVVVQSQQWAIHV 149

RESULT 12
Q8CW64 PRELIMINARY; PRT; 160 AA.
ID Q8CW64
AC Q8CW64;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RA "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.5%; Score 104.5; DB 16; Length 160;
Best Local Similarity 29.2%; Pred. No. 0.58;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAATVVGSAAL--AGVVQWGGGNGHNGSGSPDSTLSIY-QYGSANAALALQSDARK 65
DB 42 FAVNELSKSFNQAAII--GQAGTNNSAQLRQGGSKLLTVVAQEGSSNRA-KIDQTGDY 97

QY 66 SETTITQSGYNGADVGQAGNADNSTIELTQNGFNFNATIDQWNAKNSDITVGYDQLVTRV 125

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVZQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 757 | 100.0 | 151 | 3 | AAB36348 |
| 2 | 658 | 85.9 | 151 | 2 | AAR74625 |
| 3 | 658 | 85.9 | 151 | 3 | AAB36341 |
| 4 | 653 | 85.3 | 151 | 2 | AAW23570 |
| 5 | 618 | 81.6 | 151 | 3 | AAB36349 |
| 6 | 580 | 76.6 | 151 | 3 | AAB36346 |
| 7 | 578 | 76.4 | 151 | 3 | AAB36347 |
| 8 | 577 | 76.2 | 151 | 3 | AAB36353 |
| 9 | 575 | 76.0 | 151 | 3 | AAB36352 |
| 10 | 574 | 75.8 | 151 | 3 | AAB36350 |
| 11 | 568 | 75.0 | 151 | 3 | AAB36354 |
| 12 | 567 | 74.9 | 151 | 3 | AAB36351 |
| 13 | 566 | 74.8 | 151 | 3 | AAB36355 |
| 14 | 487 | 64.3 | 120 | 2 | AAR62761 |
| 15 | 487 | 64.3 | 120 | 2 | AAW23569 |
| 16 | 484 | 63.9 | 151 | 3 | AAB36343 |
| 17 | 479 | 63.3 | 151 | 7 | ABR82651 |
| 18 | 406 | 53.6 | 142 | 2 | AAR52664 |
| 19 | 345 | 45.6 | 122 | 2 | AAR52663 |
| 20 | 237 | 31.3 | 45 | 3 | AAB36316 |
| 21 | 123 | 16.2 | 23 | 3 | AAB36321 |
| 22 | 123 | 16.2 | 23 | 3 | AAB36326 |
| 23 | 123 | 16.2 | 23 | 3 | AAB36338 |
| 24 | 118.5 | 15.7 | 151 | 3 | AAB36344 |
| 25 | 115 | 15.2 | 22 | 3 | AAB36325 |

| | | | | | |
|----|-------|------|------|---|----------|
| 26 | 115 | 15.2 | 22 | 3 | AAB36339 |
| 27 | 115 | 15.2 | 22 | 3 | AAB36320 |
| 28 | 111.5 | 14.7 | 151 | 3 | AAB36342 |
| 29 | 111 | 14.7 | 22 | 3 | AAB36322 |
| 30 | 111 | 14.7 | 22 | 3 | AAB36327 |
| 31 | 111 | 14.7 | 22 | 3 | AAB36337 |
| 32 | 109 | 14.4 | 23 | 3 | AAB36340 |
| 33 | 109 | 14.4 | 23 | 3 | AAB36324 |
| 34 | 109 | 14.4 | 23 | 3 | AAB36319 |
| 35 | 102 | 13.5 | 26 | 7 | ABR82649 |
| 36 | 98 | 12.9 | 26 | 7 | ABR82645 |
| 37 | 96.5 | 12.7 | 502 | 2 | AAW32312 |
| 38 | 96.5 | 12.7 | 2016 | 6 | AAB36891 |
| 39 | 96 | 12.7 | 19 | 3 | AAB36323 |
| 40 | 96 | 12.7 | 19 | 3 | AAB36336 |
| 41 | 96 | 12.7 | 19 | 3 | AAB36328 |
| 42 | 95 | 12.5 | 24 | 7 | ABR82647 |
| 43 | 95 | 12.5 | 764 | 6 | AAB36890 |
| 44 | 93 | 12.3 | 290 | 6 | ABU44579 |
| 45 | 92 | 12.2 | 23 | 3 | AAB36331 |

ALIGNMENTS

RESULT 1

AAB36348
ID AAB36348 standard; protein; 151 AA.
XX AAB36348;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.
XX
XX Salmorella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64624.
DR
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.
PT
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 757; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.2e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine.
 XX
 OS Salmonella.
 XX
 PN WO9425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 XX 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compns. contg. fimbrial type proteins.
 PT
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 XX The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct FN field.)

XX
 SQ Sequence 151 AA;

Query Match 86.9%; Score 658; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.1e-57;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 3
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.

XX
 XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;
 Query Match 86.9%; Score 658; DB 3; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.6e-57;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGYDQVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGYDQVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4
 AAW23570

ID AAW23570 standard; protein; 151 AA.

XX AC AAW23570;

XX DT 25-MAR-2003 (revised)

XX DT 29-SEP-1997 (first entry)

XX DE Salmonella enteritidis 27655-3b agfa.

XX KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.

XX OS Salmonella enteritidis.

XX FH Key Location/Qualifiers

XX FT Misc-difference 123

XX FT /note= "Encoded by GCC"

XX PN US5635617-A.

XX PD 03-JUN-1997.

XX PF 26-APR-1994; 94US-00233788.

XX PR 26-APR-1993; 93US-00054452.

XX PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX PI Collinson SK, Kay WM, Doran JL;

XX DR WPI; 1997-309886/28.

XX DR N-PSDB; AAT74142.

XX PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.

XX PS Example 2; Fig 7; 85pp; English.

XX CC The present sequence represents agfa encoded by the full agfa gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)

XX SQ Sequence 151 AA;

Query Match 86.3%; Score 653; DB 2; Length 151;
 Best Local Similarity 89.4%; Pred. No. 5.1e-57;
 Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGYDQVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGYDQVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 5

AAAB36349

ID AAB36349 standard; protein; 151 AA.

XX AC AAB36349;

XX DT 26-FEB-2001 (first entry)

XX DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.

XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

XX KW vaccine; immune response; immunogen.

XX OS Salmonella enteritidis.

XX OS Escherichia coli.

XX OS Synthetic.

XX PN WO2000060102-A2.

XX PD 12-OCT-2000.

XX PF 05-APR-2000; 2000WO-CA000356.

XX PR 05-APR-1999; 99US-0127888P.

XX PA (UYVI-) UNIV VICTORIA.

XX PI White AP, Doran JL, Collinson SK, Kay WM;

XX DR WPI; 2000-672631/65.

XX DR N-PSDB; AAC64625.

XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

XX PS Disclosure; Page 136; 139pp; English.

XX CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 81.6%; Score 618; DB 3; Length 151;
 Best Local Similarity 76.6%; Pred. No. 1.6e-53;
 Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGV-----YDQLVTRVVTHEMAHAGS 40
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSDPDYDQLVTRVVTHEMAH--- 57
 QY 41 PSTLSIYGYGSANAALALQSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNN 100
 DB 58 -----ALQSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNN 100
 QY 101 ATIDQWAKNSDITVGOYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151
 DB 101 ATIDQWAKNSDITVGOYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151

RESULT 6
 AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 XX 26-FEB-2001 (first entry)
 XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 DE *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 PD 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 PR (UWI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 FI WPI; 2000-672631/65.
 XX N-PSDB; AAC64622.
 DR Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 FT

PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CSA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 76.6%; Score 580; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 9.8e-50;
 Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSDPSTLSIYGYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSDPSTLSIYGYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWAKNSDITVGOYGG 120
 QY 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151

RESULT 7
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX 26-FEB-2001 (first entry)
 XX Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 DE *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 XX

| | | |
|----|---|----------------|
| PR | 05-APR-1999; | 99US-0127888P. |
| XX | (UYVI-) UNIV VICTORIA. | |
| PA | | |
| XX | | |
| PI | White AP, Doxan JL, Collison SK, Kay WW; | |
| XX | | |
| DR | WPI; 2000-672631/65. | |
| XX | N-PSDB; AAC64623. | |
| DR | | |
| XX | | |
| PT | Recombinant agfa gene having a segment replaced by a foreign DNA sequence | |
| PT | which encodes foreign epitope or antigen, expresses recombinant Agfa | |
| PT | protein useful for eliciting immune response in animal. | |
| XX | | |
| XX | Disclosure; Page 136; 139pp; English. | |
| XX | | |
| CC | The present invention describes a recombinant agfa gene (I) where a | |
| CC | segment of the gene has been replaced by a segment of a foreign DNA | |
| CC | sequence which encodes a foreign epitope or antigen. Also described are: | |
| CC | (1) use of thin aggregative fimbriae (SF17/TAf) nucleation depended | |
| CC | assembly system of strains of Salmonella, Escherichia coli and | |
| CC | Enterobacteriaceae for the production of fimbriae comprising recombinant | |
| CC | Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) | |
| CC | directing recombination of a recombinant gene into the chromosome of the | |
| CC | homologous species; (3) directing recombination of a recombinant gene | |
| CC | back into the chromosome of the homologous species, replacing the native | |
| CC | copy of that gene; and (4) eliciting an immune response in an animal, | |
| CC | comprising separating an amino acid polymer comprising a recombinant Agfa | |
| CC | protein containing a replacement segment or segments of foreign amino | |
| CC | acid sequence or sequences grown on a salmonella, E. coli or | |
| CC | Enterobacteriaceae host cell, from the host cell and introducing the | |
| CC | polymer into the animal in conjunction with a carrier or diluent. (1) is | |
| CC | useful for the expression of recombinant Agfa protein which is useful for | |
| CC | eliciting an immune response in an animal. In a fimbrial presentation | |
| CC | system the heterologous antigens are presented in high numbers (up to | |
| CC | 500,000 copies/cell), the hybrid fimbrial protein possesses both the | |
| CC | immunogenicity and adhesion properties relevant for an efficient live | |
| CC | vaccine, the carrier fimbrial subunit proteins are usually strong | |
| CC | immunogens, which may be important for directing an immune response | |
| CC | against the inserted epitope, and hybrid fimbriae are easy and | |
| CC | inexpensive to purify in large amount. The present sequence is given in | |
| CC | the exemplification of the present invention | |
| XX | | |
| SQ | Sequence 151 AA; | |
| XX | | |

XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX DR N-P8DB; AAC64629.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein; useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA, and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein, containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX Sequence 151 AA;
SQ

KW salmonella; agia; chromosomal gene replacement; limbrin; epitope;
KW vaccine; immune response; immunogen.

AAB36352
 ID AAB36352 standard; protein; 151 AA.
 AC AAB36352;
 XX
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 Query Match 76.0%; Score 575; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred No. 3, 1e-49;
 Matches 123; Conservative 5; Mismatches 22; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVYDQVTRVVTTHMAHAGSDPTLSITVQYGSANAALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVYDQVTRVVTTHMAHAGSDPTLSITVQYGSANAALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNAITDQWNAKNSDITVQCYGG 120
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNAITDQWNAKNDQLVTRVVT 120
 QY 121 NNAALVNOTASDSSVVRQVYGFNGNATANQY 151
 DB 121 HEWAHANQTSADSSVVRQVGFNGNATANQY 151

RESULT 10
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX

PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 75.8%; Score 574; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. NO. 3.9e-49;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 11
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
AC AAB36354;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa: PT3#9 amino acid sequence SEQ ID NO:28.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64630.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 75.0%; Score 568; DB 3; Length 151;
Best Local Similarity 80.1%; Pred. NO. 1.5e-48;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 12
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa: PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64627.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CSgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.9%; Score 567; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.9e-48;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 13

AAB36355
 ID AAB36355 standard; protein; 151 AA.

XX AAB36355;

XX 26-FEB-2001 (first entry)

DE AgfA::P73#10 amino acid sequence SEQ ID NO:30.

XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000355.

XX 05-APR-1999; 99US-012788EP.

PA (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64631.

XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 139; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CSgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.8%; Score 566; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.4e-48;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 14

AAR62761
 ID AAR62761 standard; protein; 120 AA.

XX AAR62761;

XX 27-AUG-2003 (revised)

XX 25-MAR-2003 (revised)

XX 26-JUN-1995 (first entry)

DE AgfA sequence.

XX Salmonella; AgfA; vaccine.

XX Salmonella enteritidis.

XX WO9425598-A2.

XX 10-NOV-1994.

```
XX 26-APR-1994; 94WO-IB000207.
XX 26-APR-1993; 93US-00054452.
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING) KING J.
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI; 1994-358275/44.
XX N-PSDB; AAQ73066.
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
XX strains, vector constructs, or compens. contg. fimbrial type proteins.
XX Disclosure; Fig 7A; 95pp; English.
XX The sequence represents the Salmonella enteritis 27655-3b TnpHoA mutant
XX strain Agfa protein. The encoding DNA and isolated Agfa protein are used
XX in genetic immunization and vaccine compositions, respectively, to elicit
XX an immune response to Salmonella in animals (e.g. food producing animals)
XX and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
XX AUG-2003 to correct OS field.)
XX Sequence 120 AA;
SQ
Query Match 64.3%; Score 487; DB 2; Length 120;
Best Local Similarity 97.9%; Pred. No. 1.3e-40;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 37 HASGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 96
Db 16 NSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 75
QY 97 FRNNATIDQWNAKNSDITVQYGGNNAALVNQTSADS 133
Db 76 FRNNATIDQWNAKNSDITVQYGGNNAALVNQTSADS 112
RESULT 15
AAW23569
ID AAW23569 standard; protein; 120 AA.
XX AAW23569;
XX 25-MAR-2003 (revised)
XX 29-SEP-1997 (first entry)
XX Salmonella enteritis 27655-3b TnpHoA mutant agfa fragment.
XX Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
XX Salmonella enteritidis.
XX US5635617-A.
XX 03-JUN-1997.
XX 26-APR-1994; 94US-00233788.
XX 26-APR-1993; 93US-00054452.
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX Collinson SK, Kay WW, Doran JL;
XX WPI; 1997-309886/28.
XX N-PSDB; AAT74141.
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
```

```
PS Example 2; Fig 7; 85pp; English.
XX The present sequence represents an agfa fragment encoded by an agfa gene
XX fragment derived from Salmonella enteritidis 27655-3b TnpHoA mutant
XX strain. The nucleic acid can be used to provide diagnostic assays for
XX Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.
XX It can also be used to provide proteins and antibodies which can be used
XX for assays. The nucleic acid sequence can be used to provide probes or
XX primers which can specifically hybridise to nucleic acid molecules from
XX greater than 99% of Salmonella strains that are pathogenic to warm-
XX blooded animals relative to nucleic acid molecules from virtually all
XX other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)
SQ Sequence 120 AA;
Query Match 64.3%; Score 487; DB 2; Length 120;
Best Local Similarity 97.9%; Pred. No. 1.3e-40;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 37 HASGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 96
Db 16 NSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNG 75
QY 97 FRNNATIDQWNAKNSDITVQYGGNNAALVNQTSADS 133
Db 76 FRNNATIDQWNAKNSDITVQYGGNNAALVNQTSADS 112
Search completed: August 2, 2004, 14:48:25
Job time : 45.9 secs
```

This Page Blank (uspto)

Db 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNGSGPDSSTLSIYQYGSANAALAQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Db 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQASDSSVMVROVGFNNATANQY 151
 Db 121 NNPALVNQASDSSVMVROVGFNNATANQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouthier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; TITLE OF INVENTION: OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.3%; Score 487; DB 1; Length 120;
 Best Local Similarity 97.9%; Pred. No. 2.5e-43;
 Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQGDADNSTIELTQNG 96
 Db 16 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQGDADNSTIELTQNG 75
 QY 97 FRNNATIDQWNAKNSDITVGOYGGNNAALVNQASDS 133
 Db 76 FRNNATIDQWNAKNSDITVGOYGGNNAALVNQASDS 112

RESULT 3

US-09-543-681A-5434
 ; Sequence 5434, Application US/09543681A
 ; Patent No. 6605709
 ; GENERAL INFORMATION:

; APPLICANT: GARY BRETON
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 ; FILE REFERENCE: 2709.1002-001
 ; CURRENT APPLICATION NUMBER: US/09/543,681A
 ; CURRENT FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: US 60/128,706
 ; PRIOR FILING DATE: 1999-04-09
 ; NUMBER OF SEQ ID NOS: 8344
 ; SEQ ID NO 5434
 ; LENGTH: 2315
 ; TYPE: PRT
 ; ORGANISM: Proteus mirabilis
 ; US-09-543-681A-5434

Query Match 11.6%; Score 88; DB 4; Length 2315;
 Best Local Similarity 28.0%; Pred. No. 6;
 Matches 37; Conservative 15; Mismatches 54; Indels 26; Gaps 6;

QY 24 DQAVTRVVTHEMAHAGSPD-STLSIYQYGSANAALQSDARKSETTITQSGYNGADVG 82
 Db 1700 DVTVTTLTNDAGHAGAPDYSLAGSQH---DKIALHGDGK---VTYDGASTVTVVI 1752
 QY 83 QGADNS-----TIELTQNGFRNNATIDQWNAKNSDITVGOYGG--NNAALVNQASDS 133
 Db 1753 PAGSKSVSFIVDPTMEANQNAF-----NAEGMEKVVAITITGTSNNATAVTDIVNNA 1803
 QY 134 SVMVROVGFNN 145
 Db 1804 GASATGVYDGN 1815

RESULT 4

US-09-477-135A-131
 ; Sequence 131, Application US/09477135A
 ; Patent No. 6572865
 ; GENERAL INFORMATION:
 ; APPLICANT: Nano, Francis
 ; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
 ; TITLE OF INVENTION: immunostimulatory Peptides
 ; FILE REFERENCE: 52888
 ; CURRENT APPLICATION NUMBER: US/09/477,135A
 ; CURRENT FILING DATE: 2000-01-03
 ; PRIOR APPLICATION NUMBER: 08990823
 ; PRIOR FILING DATE: 1997-12-15
 ; PRIOR APPLICATION NUMBER: US 96/10375
 ; PRIOR FILING DATE: 1996-06-14
 ; PRIOR APPLICATION NUMBER: 60/000,254
 ; PRIOR FILING DATE: 1995-06-15
 ; NUMBER OF SEQ ID NOS: 169
 ; SOFTWARE: Patent in Ver. 2.0
 ; SEQ ID NO 131
 ; LENGTH: 943
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; US-09-477-135A-131

Query Match 11.4%; Score 86.5; DB 4; Length 943;
 Best Local Similarity 24.8%; Pred. No. 2.4;
 Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AFAAAVWSG-----SALAGVYDQVLTAVTVTHEMAHAGSPDSTLS-IYQYGSANAA 56
 Db 96 AUFVSLVSNLLGQNAPATAEAYEQMWADQVAMFGYHAGAAVSAITPFGQALPT 155
 QY 57 L-----ALQSDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKNSD 112
 Db 156 VAGGGALVSAAAAQVTTTRVFRNLGL-ANVGEQN-----VGNNGVGNFNLGSANIGNGN 207
 QY 113 ITVGOYGGNNAALVNQASDSSVMVROVGFNNATAN 149
 Db 208 IGSNGTSSNIGFGN-VGPGLTAALNNIGFGNTGSNN 243

```
RESULT 5
US-09-252-991A-26438
; Sequence 26438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26438
; LENGTH: 1415
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26438

Query Match 11.2%; Score 85; DB 4; Length 1415;
Best Local Similarity 25.6%; Pred. No. 6.1; Indels 52; Gaps 7;
Matches 43; Conservative 19; Mismatches 54;

QY 24 DQLVT---RVVTHEM-----AHASGPDSTLSIVQYGSANA-----ALA 58
DB 357 DQATGTWRLSHEVSAVQCAAGSGQVHVAQGPAGANASDSNGVTIVQQPAVD 416
QY 59 LOSDARKSETTITQSCYG-----NGADVQGGADNSTIELTON-----GFR 98
DB 417 LAAGANGTSVAVQSQSGANGISVQVQSGNANGAGASDISVQSQNSPNIGSGVN 476
QY 99 NNATIDOWNAXN----SDIVGVQYGGNNALVNOTASDSSVMVRQVG 141
DB 477 GVTVVQSQNGANGISGASGITVQV--SQNGANGISGASISVQVQSQG 522

RESULT 6
US-08-194-290-7
; Sequence 7, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; TITLE OF INVENTION: Bacterial surface protein expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shiesinger, Arkwright & Garvey
; STREET: 3000 South Eads Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garvey, George A
; REGISTRATION NUMBER: 17737
; REFERENCE/DOCKET NUMBER: 5946-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-5600
; TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 7:

Query Match 11.2%; Score 82; DB 1; Length 1026;
Best Local Similarity 24.8%; Pred. No. 7.9; Indels 20; Gaps 7;
Matches 34; Conservative 28; Mismatches 55;

QY 11 AIVVSGSALAGYVDQLVIRVTHEMAHA---SGPDSTLSIVQYGSNAALALQSDARKSE 67
DB 463 AAATAGATVAGRVNGAVT--ITDAAASATTAGKATVTLGSGAA---TIDSSAL--- 513
QY 68 TTITQSGYGNADVQGA-----DNSTIELTONGFENNATI-DQWNAKNSDITVQYGGN 121
DB 514 TTVNLSGTGSLGIRGALTATPTANTLTNVNGLTTTGATIDSEAAADGGTTINIAGS 573
QY 122 NAA--LVNQTASDSSVM 136
DB 574 TASSTIASLVAAADATTL 590

RESULT 7
US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/614,377A
; FILING DATE: 12-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 07/895,367
; FILING DATE: 03-JUNE-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsao, Y. Rocky
; REGISTRATION NUMBER: 34053
; REFERENCE/DOCKET NUMBER: 08106/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1026 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```


APPLICANT: Weltzin, Richard A.
APPLICANT: Guy, Bruno
TITLE OF INVENTION: LT and CT in Parenteral Immunization
TITLE OF INVENTION: Methods Against Helicobacter Infection
FILE REFERENCE: 06132/055002
CURRENT APPLICATION NUMBER: US/09/336,115C
CURRENT FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: US 09/100,258
PRIOR FILING DATE: 1998-06-19
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 691
TYPE: PRT
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: SIGNAL
LOCATION: (1)...(18)
US-09-336-115C-12

Query Match 10.7%; Score 81; DB 4; Length 691;
Best Local Similarity 19.0%; Pred. No. 5.7;
Matches 33; Conservative 27; Mismatches 66; Indels 48; Gaps 5;

QY 8 AFAIVSGSALAGVYDQVTRV-----VTHEMAHAGPD 42
DB 100 AQAVALNALAAVGMQVIALFIGCGPGTNNQSYSGFNTPALNGTTTCNQAYGTGN 159
QY 43 STLSTIYQYGSANAALQSDARKSETTITQSYGNGADVQGGADNSITLTQNGFR--- 98
DB 160 GILSIDEYKLNQAYQI-----TQALNQGGMPALNTTKGVNIOQTNYRTTQ 213
QY 99 NNATIDQWAKNSDITVQYGG-----NNAA-LVNOTASDSSVMVRQ 139
DB 214 NNIIEHYTENGKEIPVSYSGSFPSTIQLTYHNNAENLIQQAATIMQVLITQ 267

RESULT 12
US-09-336-115C-6
Sequence 6, Application US/09336115C
Patent No. 6576244
GENERAL INFORMATION:
APPLICANT: Weltzin, Richard A.
APPLICANT: Guy, Bruno
TITLE OF INVENTION: LT and CT in Parenteral Immunization
TITLE OF INVENTION: Methods Against Helicobacter Infection
FILE REFERENCE: 06132/055002
CURRENT APPLICATION NUMBER: US/09/336,115C
CURRENT FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: US 09/100,258
PRIOR FILING DATE: 1998-06-19
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 745
TYPE: PRT
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: SIGNAL
LOCATION: (1)...(20)
FEATURE:
NAME/KEY: VARIANT
LOCATION: 721
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match 10.7%; Score 81; DB 4; Length 745;
Best Local Similarity 23.1%; Pred. No. 6.4;
Matches 27; Conservative 18; Mismatches 68; Indels 4; Gaps 2;

QY 31 VTHEMAHAGPDSTLIYQYGSANAALQSDARKSETTITQSYGNGADVQGGADNSTI 90
DB 152 ITCNRFSTGPGKWSIDEFKLNKNEAYQIIQOALKNQSGFPFLG-GNGTKV---SVNRY 207

QY 91 ELTQNGFRNNATIDQWAKNSDITVQYGGNNAALVNQOTASDSSVMVRQVGFNNAT 147
DB 208 ECRQTADINGVYQFCARNGSSSSNGGSGSTQTTATTQDGVITTTNNKAT 264

RESULT 13
US-09-206-942-69
Sequence 69, Application US/09206942
Patent No. 6432669
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
TITLE OF INVENTION: Molecular Weight Proteins
FILE REFERENCE: 1038-861 MIS-7b
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 69
LENGTH: 1095
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-206-942-69

Query Match 10.6%; Score 80.5; DB 4; Length 1095;
Best Local Similarity 26.4%; Pred. No. 12;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SGSALAGVYDQ--LVTRVYTH---EMAHAGPDSTLIYQYGSANAALQSDARKSET 69
DB 727 TGSILGIESSGSVTLTATEGALAVNSISG--NTVVTANSGLATTLAGSTIKGTESVT 784

QY 69 TITQSGYNGADVQGGADNSTIEL-----TQNGFRNNATIDQWAKNSDITVQYGG 121
DB 785 TSSQSG-----DIGTISGTVKATESLTQNSKIKATTEANVTATGTTIGTISG 839

QY 122 NAALVNQOTASDSSVMVRQVGFNNATAN 149
DB 840 NTVVTANAGDLTV-----GNGAEIN 860

RESULT 14
US-08-038-682-2
Sequence 2, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GENE III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2

Query Match 10.6%; Score 80.5; DB 1; Length 1536;
Best Local Similarity 26.4%; Pred. No. 20;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SGSALAGVYDQ--LVTRVVTTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 1168 TGSILGGIESSSGSVTLTATGALAVSNISG--NTVTVTANSGLTTLTLAGSTIKGTESVT 1225

QY 69 TITQSGYNGADYVGQADNSTIEL-----TONGFRNNATIDOWNAKNSDITVQYQYGN 121
Db 1226 TSSQSG-----DIGGTISGTVVEVKATESLTTSQNSKIKAATTEANVTSATGTIGGTISG 1280

QY 122 NAALVNQATASDSSVMVRQVFGNNATAN 149
Db 1281 NTVNVTANAGDLTV-----GNGAEIN 1301

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-2

Query Match 10.6%; Score 80.5; DB 1; Length 1536;
Best Local Similarity 26.4%; Pred. No. 20;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

QY 15 SGSALAGVYDQ--LVTRVVTTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 1168 TGSILGGIESSSGSVTLTATGALAVSNISG--NTVTVTANSGLTTLTLAGSTIKGTESVT 1225

QY 69 TITQSGYNGADYVGQADNSTIEL-----TONGFRNNATIDOWNAKNSDITVQYQYGN 121
Db 1226 TSSQSG-----DIGGTISGTVVEVKATESLTTSQNSKIKAATTEANVTSATGTIGGTISG 1280

QY 122 NAALVNQATASDSSVMVRQVFGNNATAN 149
Db 1281 NTVNVTANAGDLTV-----GNGAEIN 1301

```

Search completed: August 2, 2004, 14:58:32
Job time : 12 secs

```

RESULT 15
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09D_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|----------------------|
| 1 | 481 | 63.5 | 151 | 12 | US-09-741-873B-4 |
| 2 | 481 | 63.5 | 151 | 12 | US-09-741-873B-4 |
| 3 | 414 | 54.7 | 131 | 12 | US-09-741-873B-2 |
| 4 | 414 | 54.7 | 131 | 12 | US-09-741-873B-2 |
| 5 | 113.5 | 15.0 | 445 | 15 | US-10-369-493-20638 |
| 6 | 93 | 12.3 | 290 | 12 | US-10-282-122A-72503 |
| 7 | 89.5 | 11.8 | 1862 | 12 | US-10-282-122A-49757 |
| 8 | 89 | 11.8 | 182 | 14 | US-10-238-075-1549 |
| 9 | 88.5 | 11.7 | 597 | 9 | US-09-793-306-146 |
| 10 | 86.5 | 11.4 | 154 | 16 | US-10-437-963-162284 |
| 11 | 86.5 | 11.4 | 943 | 9 | US-09-996-634-131 |
| 12 | 86.5 | 11.4 | 943 | 10 | US-09-997-182-131 |
| 13 | 86.5 | 11.4 | 943 | 10 | US-09-997-181-131 |
| 14 | 86.5 | 11.4 | 3300 | 12 | US-10-282-122A-64369 |
| 15 | 85 | 11.2 | 1554 | 16 | US-10-383-930-38 |

Sequence 107204, A
Sequence 49854, A
Sequence 1, Appli
Sequence 34, Appli
Sequence 1, Appli
Sequence 9411, Ap
Sequence 10, Appli
Sequence 11, Appli
Sequence 70017, A
Sequence 16, Appli
Sequence 33, Appli
Sequence 13173, A
Sequence 20096, A
Sequence 5343, Ap
Sequence 7, Appli
Sequence 109, App
Sequence 73678, A
Sequence 70586, A
Sequence 49773, A
Sequence 12, Appli
Sequence 22, Appli
Sequence 6, Appli
Sequence 58683, A
Sequence 73345, A
Sequence 54094, A
Sequence 65, Appli
Sequence 2, Appli
Sequence 63, Appli
Sequence 146426, A

US-10-437-963-107204
US-10-282-122A-49854
US-09-984-334-1
US-10-004-115A-34
US-10-327-108-1
US-10-156-761-9411
US-10-185-990-10
US-10-185-990-11
US-10-282-122A-70017
US-09-996-194-16
US-10-164-966-33
US-10-156-761-13173
US-10-369-493-20096
US-10-369-493-5343
US-09-379-931-7
US-10-223-397-7
US-09-769-787-109
US-10-282-122A-73678
US-10-282-122A-70586
US-10-282-122A-49773
US-08-834-666A-12
US-08-834-666A-22
US-08-834-666A-6
US-10-282-122A-58683
US-10-282-122A-73345
US-10-282-122A-54094
US-10-193-764-65
US-10-092-880-2
US-10-193-764-63
US-10-437-963-146426

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09/741,873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 63.5%; Score 481; DB 12; Length 151;
Best Local Similarity 66.2%; Pred. No. 2.1e-44;
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVWTHENAHASGPDSTLSIYCGSANAALQ 60

```

Db      1  MKLLKVAALAAIFVSGSAGVVPVQVGGGNGGNSGPNSELNIYQVGGNSALALQ 60
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      61  SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNATIDQWNAKNSDITVQYGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61  TDARNSDLTITQGGGNGADVQGGSDSDIDLTQGRFGNSATLDQWNGKNSMTVKQFGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      121  NNAALVNOTASDSSVMVRQVGFNNATANQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 2

```

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match      63.5%; Score 481; DB 12; Length 151;
Best Local Similarity 56.2%; Pred. No. 2.1e-44;
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

```

```

Qy      1  MKLLKVAAPAAIVSGSALAGVVDLVTRVVTHEMAHAGSDPSTLSIYQGSANAALALQ 60
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1  MKLLKVAALAAIFVSGSAGVVPVQVGGGNGGNSGPNSELNIYQVGGNSALALQ 60
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      61  SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNATIDQWNAKNSDITVQYGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61  TDARNSDLTITQGGGNGADVQGGSDSDIDLTQGRFGNSATLDQWNGKNSMTVKQFGG 120
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      121  NNAALVNOTASDSSVMVRQVGFNNATANQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 3

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      54.7%; Score 414; DB 12; Length 131;
Best Local Similarity 69.9%; Pred. No. 3.6e-37;
Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;
Qy      39  SGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVQGGADNSTIETQNGFR 98
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      19  SGENSELTYQGGNSALALQTDARNSDLTITQGGGNGADVQGGSDSDIDLTQGRFG 78
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      99  NNATIDQWNAKNSDITVQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      79  NSATLDQWNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQVGFNNATAHQY 131
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      54.7%; Score 414; DB 12; Length 131;
Best Local Similarity 69.9%; Pred. No. 3.6e-37;
Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;
Qy      39  SGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVQGGADNSTIETQNGFR 98
          :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49757
; LENGTH: 1862
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49757

Query Match      11.8%; Score 89.5; DB 12; Length 1862;
Best Local Similarity 26.8%; Pred. No. 3.7;
Matches 51; Conservative 16; Mismatches 66; Indels 57; Gaps 8;

QY      7 AFAAAT-----VSSGALAGVDQVTRVTHEMAHA--SGPDSLSIYQYGSANA 55
Db      1278 AASATATANGDVTVASAGKLAAGTTVGTVTALNDINAGATIESNGDAVLNA-QQGSLSNA 1336

QY      56 ALALQSDARKSETTITQSGYNGADVGGQADNS-----TIELTQNGFRNNATIDQW--- 107
Db      1337 TGGINGSA---ELTIT-----TGLDLSLCASTSAVGCDVTLNAGRNAILNGTLVGQNGYI 1388

QY      108 AKNSDT-----VQYGGNNAALVNQATSDSSVMVR 138
Db      1389 SAQGDVTGPTGQFTNAAVLGAQGDIALTGSQANSVQATGCDGDAALNNVTSSITLTILTA 1448

QY      139 QVGFGNATA 148
Db      1449 NGNAGNGDAA 1458

RESULT 8
US-10-238-075-1549
; Sequence 1549, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1549
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-1549

Query Match      11.8%; Score 89; DB 14; Length 182;
Best Local Similarity 24.5%; Pred. No. 0.17;
Matches 39; Conservative 29; Mismatches 71; Indels 20; Gaps 7;

```

```

QY      1 MLLKVAFAAIVVSGALAGVDQVLT---RVVTHE-MAHASGPDSTLSIYQYGSANA 56
Db      3 IKVIALATFVSAPFAGSAMA--YDGTITFTGKVVAQCTVNTSDKDLAVTL-----PTVAT 56

QY      57 LALQSDARKSETT---ITQSGYNGADVGGQA-----DNSTIELTQNGFRNNATIDQWNA 108
Db      57 SSLKNAATSGLTPTFAIRLTGCATGMNSAQNVKAYFEPSSNIDLATHNLKNTAT--PTKA 114

QY      109 KNSDITVQYGGNNAALVNQATSDSSVVRQVGFGNAT 147
Db      115 DNVQIQLNSNGSTILLGEADNGQDVQSETIGSDGSAT 153

RESULT 9
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
US-09-793-306-146

Query Match      11.7%; Score 88.5; DB 9; Length 597;
Best Local Similarity 35.9%; Pred. No. 1;
Matches 28; Conservative 7; Mismatches 32; Indels 11; Gaps 4;

QY      74 GYGNADVGGQADNSTIELTQNGFRNNATIDQ--MNAKNSDITVQYGGNNAALVNQTAS 131
Db      387 GFGNAGDINTGFCNAGD--TNTGFGNAGFFNMIGNAGNEDMGVNGGSGFNVGVGN--AG 442

QY      132 DSSVMVRQVGFGNATAN 149
Db      443 NQS-----VGFGNAGTLN 455

RESULT 10
US-10-437-963-162284
; Sequence 162284, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B

```


QY 7 AAFAAIVVSG-----SALAGVYDQVTRVVTWTHMAHAGSPDSTLS-IYQYGSANAA 56
 Db 96 ALFVSLVSNLLGQAPATAEAAEQWAOVAAAFYHAGAGAAVSALTFFGQALPT 155
 QY 57 L-----ALQSDARKSETTITQSGYNGADYVGGADNSSTIELTONGFRNNATIDOWNAKNSD 112
 Db 156 VAGGALVSAARAAQVTRVFRNLGL-ANVGEGN-----VGNNGVGNFNLGSANIGNGN 207
 QY 113 ITVQYGGNNAALVNQTSADSSVMVRQVGFNNATAN 149
 Db 208 IGSNGIGSSNIGFNG-VGPCLTAALNNIGFGNTGSNN 243

RESULT 14

US-10-282-122A-64369
 ; Sequence 64369, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Karl
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 64369
 ; LENGTH: 3300
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 US-10-282-122A-64369

Query Match 11.4%; Score 86.5; DB 12; Length 3300;
 Best Local Similarity 24.8%; Pred. No. 17;
 Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;
 QY 7 AAFAAIVVSG-----SALAGVYDQVTRVVTWTHMAHAGSPDSTLS-IYQYGSANAA 56
 Db 115 ALFVSLVSNLLGQAPATAEAAEQWAOVAAAFYHAGAGAAVSALTFFGQALPT 174
 QY 57 L-----ALQSDARKSETTITQSGYNGADYVGGADNSSTIELTONGFRNNATIDOWNAKNSD 112

Db 175 VAGGALVSAARAAQVTRVFRNLGL-ANVGEGN-----VGNNGVGNFNLGSANIGNGN 226
 QY 113 ITVQYGGNNAALVNQTSADSSVMVRQVGFNNATAN 149
 Db 227 IGSNGIGSSNIGFNG-VGPCLTAALNNIGFGNTGSNN 262
 RESULT 15
 US-10-383-930-38
 ; Sequence 38, Application US/10383930
 ; Publication No. US20040127400A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, Daniel J
 ; APPLICANT: Taubman, Martin A
 ; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
 ; FILE REFERENCE: 25669-018
 ; CURRENT APPLICATION NUMBER: US/10/383,930
 ; CURRENT FILING DATE: 2003-03-07
 ; PRIOR APPLICATION NUMBER: 60/402,483
 ; PRIOR FILING DATE: 2002-08-08
 ; PRIOR APPLICATION NUMBER: 60/363,209
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 38
 ; LENGTH: 1554
 ; TYPE: PRT
 ; ORGANISM: Streptococcus sobrinus
 US-10-383-930-38

Query Match 11.2%; Score 85; DB 16; Length 1554;
 Best Local Similarity 22.7%; Pred. No. 9;
 Matches 27; Conservative 22; Mismatches 60; Indels 10; Gaps 2;
 QY 42 DSTLSIYQYGSANAAALQSDARKSETTITQSGY-----GNGADYVGGADNSSTIEL 92
 Db 761 DSDTNIVKYTDANGNLFTTADDIKGYSTVDMGSLAVVVPVGAKGQDVRVAADTNQ-KA 819
 QY 93 TQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151
 Db 820 DGKSLKTSALDSQVIYEGFSNFQDFANNADYTNKKIAENADFFKGLGITSFENAPQY 878

Search completed: August 2, 2004, 15:36:10
 Job time : 37.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-16
Perfect score: 757
Sequence: 1 MLLKVAFAAIVVSGSALA.....DSSVMVRQVFGGNATANQV 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
 - 2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
 - 3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
 - 4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
 - 5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
 - 6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
 - 7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
 - 8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
 - 9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
 - 10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
 - 11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
 - 12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
 - 13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
 - 14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
 - 15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
 - 16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
 - 17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
 - 18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
 - 19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
 - 20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
 - 21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
 - 22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
 - 23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
 - 24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
 - 25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
 - 26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
 - 27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
 - 28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
 - 29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
 - 30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
 - 31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
 - 32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
 - 33: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|-------|-------------|
| ----- | ----- | ----- | ----- | ----- | ----- |

| | | | | | | |
|----|-------|-------|-----|----|----------------------|--------------------|
| 1 | 757 | 100.0 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 2 | 658 | 86.9 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 3 | 653 | 86.3 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 4 | 618 | 81.6 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 5 | 580 | 76.6 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 6 | 579 | 76.5 | 131 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 7 | 578 | 76.4 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 8 | 577 | 76.2 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 9 | 575 | 76.0 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 10 | 574 | 75.8 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 11 | 568 | 75.0 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 12 | 567 | 74.9 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 13 | 566 | 74.8 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 14 | 554 | 73.2 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 15 | 487 | 64.3 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 16 | 484 | 63.9 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 17 | 481 | 63.5 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 18 | 479 | 63.3 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 19 | 479 | 63.3 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 20 | 479 | 63.3 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 21 | 468 | 61.8 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 22 | 468 | 61.8 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 23 | 468 | 61.8 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 24 | 414 | 54.7 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 414 | 54.7 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 401 | 53.0 | 109 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 27 | 347 | 45.8 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 28 | 250.5 | 33.1 | 70 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 29 | 237 | 31.3 | 48 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 30 | 122.5 | 16.2 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 31 | 122.5 | 16.2 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 122.5 | 16.2 | 186 | 30 | US-10-417-886-5833 | Sequence 5833, Ap |
| 33 | 118.5 | 15.7 | 151 | 19 | US-09-543-407-8 | Sequence 8, Appl |
| 34 | 113.5 | 15.0 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 35 | 113.5 | 15.0 | 445 | 33 | US-60-360-039-20638 | Sequence 20638, A |
| 36 | 111.5 | 14.7 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 37 | 101.5 | 13.4 | 590 | 3 | US-07-965-673A-10 | Sequence 10, Appl |
| 38 | 101.5 | 13.4 | 590 | 22 | US-09-791-537-29229 | Sequence 29229, A |
| 39 | 101.5 | 13.4 | 590 | 22 | US-09-791-537-111372 | Sequence 111372, A |
| 40 | 101 | 13.3 | 599 | 22 | US-09-791-537-29228 | Sequence 29228, A |
| 41 | 101 | 13.3 | 599 | 22 | US-09-791-537-111371 | Sequence 111371, A |
| 42 | 99.5 | 13.1 | 582 | 22 | US-09-791-537-94843 | Sequence 94843, A |
| 43 | 99.5 | 13.1 | 646 | 22 | US-09-791-537-30765 | Sequence 30765, A |
| 44 | 99.5 | 13.1 | 646 | 22 | US-09-791-537-32862 | Sequence 32862, A |
| 45 | 98 | 12.9 | 145 | 21 | US-09-739-449-8854 | Sequence 8854, Ap |

ALIGNMENTS

RESULT 1
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIRMAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b af9a
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-16

| | Query Match | 100.0%; | Score 757; | DB 19; | Length 151; |
|----|-----------------------|---|---------------------|-----------|-------------|
| | Best Local Similarity | 100.0%; | Pred. No. 3.4e-76; | | |
| | Matches 151; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | MLLLKVAAPAAALVWSGSALAGVYDQLVTRVVTTHMAHASC | PDSTLSIYYGSGNAALALQ | 60 | |
| Db | 1 | MLLLKVAAPAAALVWSGSALAGVYDQLVTRVVTTHMAHASC | PDSTLSIYYGSGNAALALQ | 60 | |
| Qy | 61 | SDARKSETTITQSGYNGADYQGQADNSTIELTQNGFRNNATID | OWNAKNSDITVGYGG | 120 | |
| Db | 61 | SDARKSETTITQSGYNGADYQGQADNSTIELTQNGFRNNATID | OWNAKNSDITVGYGG | 120 | |
| Qy | 121 | NNAALVNOTASDSSVMVRFQVGFGNATANQY | 151 | | |
| Db | 121 | NNAALVNOTASDSSVMVRFQVGFGNATANQY | 151 | | |

RESULT 2

```

US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

```

RESULT 3

```

RESOL: 3
US-08-233-642A-57
; Sequence 57, Application US/082333642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
;
; CURRENT FILLING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18
Query Match 81.6%; Score 618; DB 19; Length 151;
Best Local Similarity 76.6%; Pred. No. 1.4e-60;
Matches 131; Conservative 0; Mismatches 0; Indels 40;

```



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match          76.2%; Score 577; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 5.7e-56;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALYDQ 60
   |||||

QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||
Db 61 LVTRVVTHEMAHAGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||

RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match          76.0%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 9.5e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALALQ 60
   |||||

QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||
Db 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
Db 121 HEWAHANQTASDSSVMVRQVGFNNATANQY 151
   |||||

RESULT 10
US-09-543-407-20
; Sequence 20, Application US/09543407
```

```
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match          75.8%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALALQ 60
   |||||

QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||
Db 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match          75.0%; Score 568; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 5.8e-55;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQYGSANAALALQ 60
   |||||

QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
   :|||
```


Db 61 SDARKSETTITQSGYGNADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDIIVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12

US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-22

Query Match 74.9%; Score 567; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 7.5e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNADYDQGVGADNVDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 120
Db 61 SDARKSETTITQSGYGNADYDQGVGADNVDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30

US-09-543-407-30

Query Match 74.8%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 9.7e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNADYDQGVGADNVDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 120
Db 61 SDARKSETTITQSGYGNADYDQGVGADNVDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14

US-09-543-407-34
; Sequence 34, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-34

Query Match 73.2%; Score 554; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 STLSIYQGSANAALALQSDARKSETTITQSGYGNADYDQGVGADNVDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 102
Db 1 STLSIYQGSANAALALQSDARKSETTITQSGYGNADYDQGVGADNVDQLVTRVVTTHMAHADQWNAKNSDIIVGOYGG 102
QY 103 IDQWNAKNSDIIVGOYGNALVNQTSADSSVMVRQVGFNNATANQY 151
Db 61 IDQWNAKNSDIIVGOYGNALVNQTSADSSVMVRQVGFNNATANQY 109

RESULT 15

US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 28-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-55

Query Match          64.3%; Score 487; DB 6; Length 120;
Best Local Similarity 97.9%; Pred. No. 5.2e-46;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIQYGSANAALALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNG 96
Db 16 NSSGPDSTLSIQYGSANAALALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNG 75

QY 97 FRNNATIDOWNAKNSDITVGYGGNNAALVNOTASDS 133
Db 76 FRNNATIDOWNAKNSDITVGYGGNNAALVNOTASDS 112

Search completed: August 2, 2004, 15:26:43
Job time : 168.9 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-16
Perfect score: 757
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVWVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.psp.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.psp.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.psp.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.psp.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.psp.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.psp.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.psp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query % | Score | Match | Length | DB | ID | Description |
|------------|---------|-------|-------|--------|----------------------|----|--------------------|
| 1 | 481 | 63.5 | 151 | 5 | US-09-741-873C-4 | | Sequence 4, Appli |
| 2 | 414 | 58.7 | 131 | 5 | US-09-741-873C-2 | | Sequence 2, Appli |
| 3 | 93 | 12.3 | 348 | 6 | US-10-004-115B-34 | | Sequence 34, Appl |
| 4 | 84 | 11.1 | 385 | 7 | US-60-556-841-1453 | | Sequence 1453, Ap |
| 5 | 82.5 | 10.9 | 597 | 6 | US-10-784-592-35 | | Sequence 35, Appl |
| 6 | 82 | 10.8 | 1236 | 6 | US-10-873-528-109 | | Sequence 109, App |
| 7 | 81.5 | 10.8 | 1158 | 6 | US-10-170-205E-32312 | | Sequence 32312, A |
| 8 | 81.5 | 10.8 | 1158 | 6 | US-10-501-035-366 | | Sequence 366, App |
| 9 | 80.5 | 10.6 | 1095 | 5 | US-09-806-709-69 | | Sequence 69, Appl |
| 10 | 80.5 | 10.6 | 1536 | 5 | US-09-806-709-69 | | Sequence 69, Appl |
| 11 | 80 | 10.6 | 485 | 6 | US-10-170-205E-11455 | | Sequence 11455, A |
| 12 | 80 | 10.6 | 503 | 6 | US-10-170-205E-11130 | | Sequence 11130, A |
| 13 | 80 | 10.6 | 503 | 6 | US-10-170-205E-11456 | | Sequence 11456, A |
| 14 | 80 | 10.6 | 546 | 6 | US-10-425-115-362263 | | Sequence 362263, A |
| 15 | 80 | 10.6 | 829 | 7 | US-60-579-902-7335 | | Sequence 7335, Ap |
| 16 | 79 | 10.4 | 4326 | 1 | PCT-US04-12717-18 | | Sequence 18, Appl |
| 17 | 79 | 10.4 | 4326 | 1 | PCT-US04-12717-18 | | Sequence 18, Appl |
| 18 | 78.5 | 10.4 | 358 | 6 | US-10-767-701-40444 | | Sequence 4044, A |
| 19 | 77 | 10.2 | 841 | 7 | US-60-565-632-7906 | | Sequence 7906, Ap |
| 20 | 77 | 10.2 | 841 | 7 | US-60-579-062-7906 | | Sequence 7906, Ap |
| 21 | 77 | 10.2 | 10203 | 6 | US-10-724-972A-4098 | | Sequence 4098, Ap |
| 22 | 76.5 | 10.1 | 388 | 6 | US-10-805-394-5044 | | Sequence 5044, Ap |
| 23 | 76.5 | 10.1 | 630 | 7 | US-60-581-351-9806 | | Sequence 9806, Ap |
| 24 | 76.5 | 10.1 | 631 | 7 | US-60-581-351-9807 | | Sequence 9807, Ap |
| 25 | 76.5 | 10.1 | 892 | 5 | US-09-952-267B-5 | | Sequence 5, Appli |
| 26 | 76.5 | 10.1 | 892 | 6 | US-10-872-768-5 | | Sequence 5, Appli |

| | | | | | | |
|----|------|------|------|---|----------------------|--------------------|
| 27 | 76.5 | 10.1 | 892 | 6 | US-10-872-769-5 | Sequence 5, Appli |
| 28 | 76.5 | 10.1 | 1627 | 7 | US-60-546-745-52 | Sequence 52, Appl |
| 29 | 76 | 10.0 | 412 | 7 | US-60-565-632-7905 | Sequence 7905, Ap |
| 30 | 76 | 10.0 | 412 | 7 | US-60-579-062-7905 | Sequence 7905, Ap |
| 31 | 76 | 10.0 | 852 | 6 | US-10-170-205E-3413 | Sequence 3413, Ap |
| 32 | 75.5 | 10.0 | 408 | 5 | US-09-248-796A-19565 | Sequence 19565, A |
| 33 | 75.5 | 10.0 | 839 | 1 | PCT-US04-12070-2 | Sequence 2, Appli |
| 34 | 75.5 | 10.0 | 1531 | 1 | PCT-US03-32645-44 | Sequence 44, Appl |
| 35 | 75.5 | 10.0 | 1924 | 1 | PCT-US04-12717-78 | Sequence 78, Appl |
| 36 | 75.5 | 10.0 | 1924 | 6 | US-10-831-070-78 | Sequence 78, Appl |
| 37 | 75 | 9.9 | 283 | 7 | US-60-565-632-12923 | Sequence 12923, A |
| 38 | 75 | 9.9 | 283 | 7 | US-60-579-062-12923 | Sequence 12923, A |
| 39 | 75 | 9.9 | 3132 | 7 | US-60-556-841-1245 | Sequence 1245, Ap |
| 40 | 75 | 9.9 | 3489 | 1 | PCT-US03-24982A-255 | Sequence 255, App |
| 41 | 75 | 9.9 | 3604 | 6 | US-10-745-237-24 | Sequence 24, Appl |
| 42 | 74.5 | 9.8 | 215 | 5 | US-09-248-796A-20202 | Sequence 20202, A |
| 43 | 74.5 | 9.8 | 237 | 6 | US-10-425-115-218336 | Sequence 218336, A |
| 44 | 74.5 | 9.8 | 393 | 6 | US-10-170-205E-6541 | Sequence 6541, Ap |
| 45 | 74.5 | 9.8 | 444 | 6 | US-10-425-115-330283 | Sequence 330283, A |

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4: Application US/09741873C
; GENERAL INFORMATION: Staffan
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 63.5%; Score 481; DB 5; Length 151;
Best Local Similarity 66.2%; Pred. No. 1.4e-39;
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | i | MKLLKVAFAAIVVSGSALAGYDQLVTRVTHMAHAGSPDSTLSIYQYGSANALALQ | 60 |
| Db | i | MKLLKVAFAAIVVSGSALAGYDQLVTRVTHMAHAGSPDSTLSIYQYGSANALALQ | 60 |
| QY | 61 | SDARKSETTITGSGYNGADVGQADNISTELTQNGFRNNATIDOWNAKNSDITVQYQYG | 120 |
| Db | 61 | TDARNSDLITTHGGGNGADVGQSDSSIDITQRCFGNSATLDQWNGKNSMTVKQFGG | 120 |
| QY | 121 | NNAALVNQASDSSVWVROVGFNNATANQY | 151 |
| Db | 121 | NGCAAVDQTASNSVNVTVQFGNNATANQY | 151 |

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-2

Query Match      54.7%; Score 414; DB 5; Length 131;
Best Local Similarity 69.9%; Pred. No. 4.3e-33;
Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;

QY 39 SPGDSLTSLYQGSANAALALQSDARKSETTTTQSGYNGADVGGCAGNDSITLTONGR 98
Db 19 SCFNSLNTYQYGGGNSALALQTDARNSLTTTQHGNGADVGGGSDSSIDLTRQGF 78
QY 99 NNATIDQWNAKSDITVGYGGNNAALVNOTASDSSVMVQVGFGNNTATNOY 151
Db 79 NSATLDQWNAKSEMIVKFGGNGRAVDQTASNSVNTVQVGFGNNTATNOY 131

RESULT 3
US-10-004-115B-34
; Sequence 34, Application US/10004115B
; GENERAL INFORMATION:
; APPLICANT: ASAKO, HIROYUKI
; APPLICANT: MATSUMURA, KENJI
; APPLICANT: SHIMIZU, MASATOSHI
; APPLICANT: ITO, NOBUYA
; APPLICANT: WAKITA, RYUHEI
; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE
; FILE REFERENCE: 7372-72249
; CURRENT APPLICATION NUMBER: US/10/004,115B
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: JP 2000-372704
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: JP 2001-006144
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 2001-026594
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: JP 2001-175175
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 34
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Corynebacterium sp.

US-10-004-115B-34
; Sequence 35, Application US/10784592
; GENERAL INFORMATION:
; APPLICANT: Wiltling, Reinhard
; APPLICANT: Ostergaard, Peter
; APPLICANT: Lassen, Soren
; TITLE OF INVENTION: POLYPEPTIDES OF ALICYCLOBACILLUS SP.
; FILE REFERENCE: 10406.203-US
; CURRENT APPLICATION NUMBER: US/10/784,592
; CURRENT FILING DATE: 2004-02-23
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Alicyclobacillus sp.
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(49)
; FEATURE:
; NAME/KEY: mat peptide
; LOCATION: (50)..(597)
; OTHER INFORMATION: multi copper oxidase
; FEATURE:

```


Db 297 RLAAARLQNGQKDVGSY-----QLPKGWSHLNGQARTSSSKLWASKSSATVFQNPNG 351
QY 65 KSETTITQSGYNGAD---VGOGADNS-----TIELTQNGFRNNATIDQWAKNSD 112
Db 352 IPSPKNGAGYKNSLERNIKQANNSLLHLKLSQTIKPMNGHSHSRGSIFFESSTP 411
QY 113 ITVQYGGNNAALVNQTSADSS 134
Db 412 TTIDEYDNNPSTDDSGDES 433

RESULT 9

US-09-806-709-69
; Sequence 69, Application US/09806709
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; FILE REFERENCE: 1038-1138 MIS
; CURRENT APPLICATION NUMBER: US/09/806,709
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/CA99/00938
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 09/206,942
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-806-709-69

Query Match 10.6%; Score 80.5; DB 5; Length 1095;
Best Local Similarity 26.4%; Pred. No. 30;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;
QY 15 SGSALAGVYDQ--LVTRVTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 727 TGIILGIESSSGSVLTATEGALAVSNISG--NTVTANSALTTLGSIKGTESVT 784
QY 69 TITQSGYNGADVQOGADNSTIEL-----TQNGFRNNATIDQWAKNSDITVQYGGN 121
Db 785 TSSQSG-----DIGGTISGGTVEVKATESLTQSNKIKATTGEANVTSGTIGGTISG 839
QY 122 NAALVNQTSADSSVMVRQVFGNNATAN 149
Db 840 NTNVNTANAGDLTV-----GNGAEIN 860

RESULT 10

US-09-806-709-67
; Sequence 67, Application US/09806709
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; FILE REFERENCE: 1038-1138 MIS
; CURRENT APPLICATION NUMBER: US/09/806,709
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: PCT/CA99/00938
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: 09/167,568
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 09/206,942
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-806-709-67

Query Match 10.6%; Score 80.5; DB 5; Length 1536;
Best Local Similarity 26.4%; Pred. No. 46;
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;
QY 15 SGSALAGVYDQ--LVTRVTH---EMAHASGPDSTLSIYQGSANAALALQS-DARKSET 68
Db 1168 TGIILGIESSSGSVLTATEGALAVSNISG--NTVTANSALTTLGSIKGTESVT 1225
QY 69 TITQSGYNGADVQOGADNSTIEL-----TQNGFRNNATIDQWAKNSDITVQYGGN 121
Db 1226 TSSQSG-----DIGGTISGGTVEVKATESLTQSNKIKATTGEANVTSGTIGGTISG 1280
QY 122 NAALVNQTSADSSVMVRQVFGNNATAN 149
Db 1281 NTNVNTANAGDLTV-----GNGAEIN 1301

RESULT 11

US-10-170-205E-11455
; Sequence 11455, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11455
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11455

Query Match 10.6%; Score 80; DB 6; Length 485;
Best Local Similarity 28.3%; Pred. No. 12;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;
QY 32 THEMAHA-----SGPDSTLSIYQGSANAALALQS-DARKSETTIT--QSGYNGA 79
Db 276 THSIAHLVPSQTVVQTFSPNPDGTVSLIQVGT--GATVATLADASELPTTIVTAQVNTSVA 334
QY 80 DVQOGADN-----STIELTQNGFRNNATIDQWAKNSDITVQYGGNNAALVNQTS 131
Db 335 D-GEVEQNWATLOGGENTICTTQASEATQAVASAEAAVAASQEMOQGGATVTTALNSEAA 393
QY 132 DSSV 135
Db 394 AHAV 397

RESULT 12

US-10-170-205E-11130
; Sequence 11130, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11130
; LENGTH: 503

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11130

Query Match      10.6%; Score 80; DB 6; Length 503;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

QY 32 THEMABA-----SGPSTLSIYQGSANAALALQSDARKSETTIT--QSGYNGGA 79
Db 294 THSIAHLVPSQTVVQTFNSNPDGTSLIQVGT-GATVATLADASELPPTVTVAQVNSAVA 352
QY 80 DVGQGAN-----STIELTQNGFRNNATIDQWNAKNSDITVGYGNGNAALVNQTAS 131
Db 353 D-GEVEQWATLQGGEMTIQTQASEATCAVASIAEAAVAASQEMQCGATVTMALNSEAA 411
QY 132 DSSV 135
Db 412 AHAV 415

RESULT 13
US-10-170-205E-111456
; Sequence 11456, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11456
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-111456

Query Match      10.6%; Score 80; DB 6; Length 503;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

QY 32 THEMABA-----SGPSTLSIYQGSANAALALQSDARKSETTIT--QSGYNGGA 79
Db 294 THSIAHLVPSQTVVQTFNSNPDGTSLIQVGT-GATVATLADASELPPTVTVAQVNSAVA 352
QY 80 DVGQGAN-----STIELTQNGFRNNATIDQWNAKNSDITVGYGNGNAALVNQTAS 131
Db 353 D-GEVEQWATLQGGEMTIQTQASEATCAVASIAEAAVAASQEMQCGATVTMALNSEAA 411
QY 132 DSSV 135
Db 412 AHAV 415

RESULT 14
US-10-425-115-362763
; Sequence 362763, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 362763
; LENGTH: 546
; TYPE: PRT
```

```
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(546)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_94013C.1.pep
US-10-425-115-362763

Query Match      10.6%; Score 80; DB 6; Length 546;
Best Local Similarity 27.2%; Pred. No. 14;
Matches 34; Conservative 14; Mismatches 43; Indels 34; Gaps 5;

QY 1 MKLLKVAAPAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLS-IYQGSANAALAL 59
Db 90 LMLRLGAVPTIVVSSPRAA-----EATR--TDHVFAFRPTSLDEILYGSDFAFSP 143
QY 60 QSD-----ARKSETTITQSGYNGADYGGQGANSTIELTON 95
Db 144 YGEHWFQARKLVTAHLFTVKRVHSYRRARKEEVRLVVAKVREAAVAGTATDMS---LAMN 200
QY 96 GPRNN 100
Db 201 TFAND 205

RESULT 15
US-60-579-902-7335
; Sequence 7335, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; FILE REFERENCE: 38-21 (53444) A
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 7335
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-7335

Query Match      10.6%; Score 80; DB 7; Length 829;
Best Local Similarity 25.2%; Pred. No. 24;
Matches 36; Conservative 24; Mismatches 53; Indels 30; Gaps 6;

QY 23 YDQLVTRVVTHEMAHAGSPDSTLSIYQGSANAALALQSD-ARKSETTITQSGY-NGAD 80
Db 474 HDEL-----SYQRKHKRFQLNLSQNLSDYGSIVYSGYLQDDWGSRTSLNLYGSVNYAD 528
QY 81 VGGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGY-----GGNNA 124
Db 529 I-----NYTLNYTONMSSNGNSTADNLAFNIQVPLDRWLPNSWATYSLNHNRCQDNQSA 583
QY 125 LVNQATASDSVM--VRQVGFNN 145
Db 584 GLSGTALDGNVGLGYLQORYGNN 606
```

Search completed: August 2, 2004, 15:29:52
Job time : 18.8 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 658 | 86.9 | 151 | 2 | JC6039 |
| 2 | 658 | 86.9 | 151 | 2 | A10635 |
| 3 | 484 | 63.9 | 151 | 2 | S70788 |
| 4 | 474.5 | 62.7 | 152 | 2 | D90806 |
| 5 | 474.5 | 62.7 | 152 | 2 | H85665 |
| 6 | 118.5 | 15.7 | 151 | 2 | S70787 |
| 7 | 118.5 | 15.7 | 151 | 2 | C90806 |
| 8 | 118.5 | 15.7 | 151 | 2 | G85665 |
| 9 | 111.5 | 14.7 | 151 | 2 | JC6040 |
| 10 | 111.5 | 14.7 | 151 | 2 | AH0635 |
| 11 | 101.5 | 13.4 | 590 | 1 | A45621 |
| 12 | 101 | 13.3 | 599 | 2 | A42049 |
| 13 | 101 | 13.3 | 599 | 2 | A4951 |
| 14 | 99.5 | 13.1 | 582 | 2 | F70675 |
| 15 | 98.5 | 13.1 | 646 | 1 | S19916 |
| 16 | 98 | 12.9 | 145 | 2 | A3143 |
| 17 | 98 | 12.9 | 145 | 2 | H38144 |
| 18 | 97.5 | 12.9 | 1034 | 1 | JC2143 |
| 19 | 96.5 | 12.7 | 602 | 1 | PL0221 |
| 20 | 96 | 12.7 | 1651 | 2 | JC1340 |
| 21 | 95.5 | 12.6 | 1258 | 2 | J00188 |
| 22 | 94.5 | 12.5 | 1322 | 2 | S7053 |
| 23 | 93.5 | 12.4 | 1567 | 2 | S11672 |
| 24 | 91.5 | 12.1 | 1655 | 2 | E97835 |
| 25 | 89.5 | 11.8 | 552 | 2 | D70604 |
| 26 | 89.5 | 11.8 | 539 | 2 | C42049 |
| 27 | 88.5 | 11.7 | 590 | 2 | E70946 |
| 28 | 87 | 11.5 | 329 | 2 | S23247 |
| 29 | 87 | 11.5 | 331 | 2 | S21406 |

30 87 11.5 331 2 S21408
31 87 11.5 455 2 C29349
32 87 11.5 573 2 C86266
33 86.5 11.4 3300 2 D70575
34 85.5 11.3 1210 2 A25547
35 84 11.1 823 2 S14055
36 84 11.1 3716 2 E70969
37 83.5 11.0 1200 1 SNPSO
38 82 10.9 528 2 S69589
39 82 10.8 534 2 T32020
40 82 10.8 1026 2 A48995
41 82 10.8 1073 2 C87374
42 82 10.8 1635 2 A10452
43 82 10.8 4776 2 E95206
44 81.5 10.8 760 2 C90739
45 81.5 10.8 760 2 E85589

outer membrane pro
hypotheical prote
F3F19.21 protein -
probable PPE prote
ice nucleation pro
nucleoskeletal-lik
probable PPE prote
ice nucleation pro
hypotheical prote
paracrystalline su
S-layer protein Rs
hemolysin (importe
cell wall surface
hypotheical prote
hypotheical prote

ALIGNMENTS

RESULT 1

JC6039
fimbriin protein agfa precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6039; PC6015; A44898
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 176, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6039
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714
A:Accession: PC6015
A:Molecule type: protein
A:Residues: 21-52 <CO2>
A:Experimental source: strain 27655-3b
A:Note: The authors translated the codon ACG for residue 44 as Ile
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella
A:Reference number: A44898; MUID:91310586; PMID:1677357
A:Contents: 27655
A:Accession: A44898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 21-33 <CO3>
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)
C:Genetics:
C:Gene: agfa
C:Function:
A:Description: major component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: fimbriin protein agfa #status experimental <MAT>

Query Match 86.9%; Score 658; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 7 5e-52;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGYVDQLVTRVTHMAHAGSPDSTLSIYGYGSANAALAQ 60
Db 1 MKLLKVAFAAIVVSGSALAGYVDQLVTRVTHMAHAGSPDSTLSIYGYGSANAALAQ 60
QY 61 SPARKSETTITGSGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SPARKSETTITGSGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Mouton, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parkhill, J.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Mouton, S.; O'Garra, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 86.9%; Score 658; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 7.5e-52;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

Db 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Y. Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42; 44-50 <OLS2>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RQRDSGLWV' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 63.9%; Score 484; DB 2; Length 151;

Best Local Similarity 66.3%; Pred. No. 2.7e-36;

Matches 101; Conservative 18; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

Db 61 TDARNSDLTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

Db 121 GNGHVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 52.7%; Score 474.5; DB 2; Length 152;

Best Local Similarity 55.8%; Pred. No. 2e-35;

Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALAL 59

Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQGSANAALAL 60

Qy 60 QSDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 119

Db 61 QADARNSLDITIQHGGNGADVGGQSDSSIDLTLQRFNGSATLDQWNGKDSHMTVKQFG 120
 QY 120 GNNALVNOTASDSSVMVROVGFNGNATANQY 151
 Db 121 GNGAAVQDTASNTVNTVQVGFNGNATANQY 152
 RESULT 5
 H5665
 Hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H5665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H5665
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-152 <STO>
 A:Cross-references: GB:AE005174; NID:G12514574; PIDN:AA655788.1; GSPDB:GN00145; UWGP:Z16
 A:Experimental source: Strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: csgA
 Query Match 62.7%; Score 474.5; DB 2; Length 152;
 Best Local Similarity 65.8%; Pred. No. 2e-35;
 Matches 100; Conservative 19; Mismatches 33; Indels 1; Gaps 1;
 QY 1 MKLLKVAFAIVGSGSALAGVYDQVTRVTH-EMAHASGPDSTLSIYQYGSANAALAL 59
 Db 1 MKLLKVAFAIAIVGSGSALAGVYQYGGGNGHGGGNSGPNSELATYQYGGGNSALAL 60
 QY 60 QSDARKSETTITQSGYNGADVGGQADNSTTIELTQGFNRNATIDQWNAKNSDITVGYG 119
 Db 61 QADARNSLDITIQHGGNGADVGGQSDSSIDLTLQRFNGSATLDQWNGKDSHMTVKQFG 120
 QY 120 GNNALVNOTASDSSVMVROVGFNGNATANQY 151
 Db 121 GNGAAVQDTASNTVNTVQVGFNGNATANQY 152
 RESULT 6
 S70787
 curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
 N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
 C:Species: Escherichia coli
 C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C:Accession: S70787; F64846
 R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A:Title: Expression of two csg operons is required for production of fibronectin- and Co
 A:Reference number: S70783; MUID:96414468; PMID:8817489
 A:Accession: S70787
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-151 <HAM>
 A:Cross-references: EMBL:X90754; NID:G1147559; PIDN:CAA62281.1; PID:G1147563
 A:Experimental source: strain K12, substrain W3110
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shaoh, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: F64846
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-151 <BLAT>
 A:Cross-references: GB:AE000205; GB:U000096; NID:G1787265; PIDN:AAC74125.1; PID:G1787278;

A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: csgB
 A:Map position: 23.15
 C:Function:
 A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr
 A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-151/Product: minor curlin chain #status predicted <MAT>
 Query Match 15.7%; Score 118.5; DB 2; Length 151;
 Best Local Similarity 31.8%; Pred. No. 0.0014;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;
 QY 35 MAHASGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQ 94
 Db 18 IAAAGYDLANSEYNF----AVNELSKSFNQAAIIQAGTNNSAQLROGGSKLLAVVAQ 73
 QY 95 NGRNNATIDQWNAKNSDITVGYGNNALVNQTASDSSVMVROVGFNGNATANQY 151
 Db 74 EGSSNRKIDQTDGYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129
 RESULT 7
 C90806
 minor curlin, subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subs
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C90806
 R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 Nat. Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90806
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA34842.1; PID:G13360879; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 050952
 C:Genetics:
 A:Gene: ECG1419
 Query Match 15.7%; Score 118.5; DB 2; Length 151;
 Best Local Similarity 31.8%; Pred. No. 0.0014;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;
 QY 35 MAHASGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQ 94
 Db 18 IAAAGYDLANSEYNF----AVNELSKSFNQAAIIQAGTNNSAQLROGGSKLLAVVAQ 73
 QY 95 NGRNNATIDQWNAKNSDITVGYGNNALVNQTASDSSVMVROVGFNGNATANQY 151
 Db 74 EGSSNRKIDQTDGYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129
 RESULT 8
 G85665
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: G85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: G85665
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-151 <STO>
 A:Cross-references: GB:AE005174; NID:G12514573; PIDN:AA655787.1; GSPDB:GN00145; UWGP:Z1

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: *csbB*

Query Match 15.7%; Score 118.5; DB 2; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0014;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy 35 MAHASGPDTLSIYQGSANAALQSDARKSETTITQSGYNGADVQ-GADNSTIETLQ 94
Db 18 IIAAGYDLANEYFN----AVNELSKSFNQAAIIGQGTNNSQLQGGSKLLAVQA 73
Qy 95 NGRFNATIDWNKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 74 EGSNNRAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNTAMIIQKSGNKANITQY 129

RESULT 9

JC6040

fimbriin protein agfB precursor - *Salmonella enteritidis*

C;Species: *Salmonella enteritidis*

C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C;Accession: JC6040

R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A;Title: *Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae.

A;Reference number: JC6039; MUID:96146512; PMID:8550497

A;Accession: JC6040

A;Molecule type: DNA

A;Residues: 1-151 <COL>

A;Cross-references: GB:U43280; NID:gl184712; PIDN:ARC43598.1; PID:gl184713

A;Experimental source: strain 276755-3b

C;Genetics:

A;Gene: *agfB*

C;Function:

A;Description: minor component of thin aggregative fimbriae

A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C;Keywords: fimbria

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-151/Product: fimbriin protein agfB #status predicted <MAT>

Query Match 14.7%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0057;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

Qy 51 GSANAALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94
Db 14 GAPGIATATNYDLARSEYNFAVNELSKSFNQAAIIGQGTNSARVRQEGSKLLSVISQ 73

Qy 95 NGRFNATIDWNKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 74 EGSNNRAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNTAMIIQKSGNKANITQY 129

RESULT 10

AH0635

nucleation component of curlin monomers [imported] - *Salmonella enterica* subsp. *enterica*

C;Species: *Salmonella enterica* subsp. *enterica* serovar Typhi

A;Note: this species has also been called *Salmonella typhi*

C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C;Accession: AH0635

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moulle, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov

A;Reference number: AH0502; MUID:21534947; PMID:11677608

A;Accession: AH0635

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-151 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gi6502314; GSPDB:GN00176

C;Genetics:

A;Gene: *STY1180*

Query Match 14.7%; Score 111.5; DB 2; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0057;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

Qy 51 GSANAALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94
Db 14 GAPGIATATNYDLARSEYNFAVNELSKSFNQAAIIGQGTNSARVRQEGSKLLSVISQ 73

Qy 95 NGRFNATIDWNKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 74 EGSNNRAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNTAMIIQKSGNKANITQY 129

RESULT 11

A45621

Leishmanolysin (EC 3.4.24.36) precursor - *Leishmania donovani*

N;Alternate names: surface endopeptidase glycoprotein gp63

C;Species: *Leishmania donovani*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jan-2000

C;Accession: A45621

R;Webb, J.R.; Button, L.L.; McMaster, W.R.

Mol. Biochem. Parasitol. 48, 173-184, 1991

A;Title: Heterogeneity of the genes encoding the major surface glycoprotein of *Leishmania*

A;Reference number: A45621; MUID:92107220; PMID:1762629

A;Accession: A45621

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-590 <WEB>

A;Experimental source: LV9

A;Note: sequence extracted from NCBI backbone (NCBIN:74958, NCBI:74959)

C;Function:

A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residues

A;Note: the activated form can activate the proenzyme form

C;Superfamily: leishmanolysin

C;Keywords: blocked carboxyl end; cell adhesion; glycoprotein; hydrolase; lipoprotein; t

F;1-39/Domain: signal sequence #status predicted <AMP>

F;40-87/Domain: activation peptide #status predicted <MAT>

F;88-565/Product: leishmanolysin #status predicted <CTP>

F;566-590/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F;48-251,255,321/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status

F;112-129,178-217,301-373,380-443,393-412,402-477,454-498,503-553,523-546/Disulfide bond

F;251,255,321/Binding site: zinc, catalytic (His) (active) #status predicted

F;252/Active site: Glu #status predicted

F;287/Binding site: carboxylate (Asn) (covalent) #status predicted

F;565/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form

Query Match 13.4%; Score 101.5; DB 1; Length 590;
Best Local Similarity 28.8%; Pred. No. 0.24;
Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;

Qy 8 AFPAIVVGSALAGVYDOLVTRVVTTHMAHASGPDSTLSIYQGSANAALQSDARK-- 65
Db 226 AVGVINIPAAIASRDQLVTRVVTTHMAHALG-----FVVFPDRARILESINVRHKDF 281

Qy 66 -----SETTITQS--GYNGA-----DVQCGADNSTIELTQNGFRNATIDQWNAK 109
Db 282 DVPVINSSTAVAKARQYGGTGLEYLEMEDQGGAGSAGSHIKM-----RNAQ-DELMAP 334

Qy 110 NSDITVQYGGNNAALVNQTA-----SDSSVMVRQVGFNNATA 146

Db 335 ASD--AGYYSALTMTAIFQDLGLFYQADF--KAEEMPWGRNA 371

RESULT 12

B42049

Leishmanolysin (EC 3.4.24.36) precursor, stationary phase - *Leishmania chagasi*

N;Alternate names: glycoprotein gp63, stationary phase; surface metalloproteinase, stat

C;Species: *Leishmania chagasi*

C;Date: 10-Mar-1994 #sequence_revision 07-Apr-1994 #text_change 22-Jun-1999

C;Accession: B42049

F;122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bond
F;261,265,331/Binding site: zinc, catalytic (His) (active) #status Predicted
F;262/Active site: Glu #status Predicted
F;394/Binding site: carboxylate (Asn) (covalent) #status predicted
F;574/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form)

Query Match 13.3%; Score 101; DB 2; Length 599;
Best Local Similarity 30.2%; Pred. No. 0.27;
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;

QY 8 AFAAIVVSGSALAGVDQLVTRVTTHMAHASPDTLSIYQYGSAHALAQSDAR--- 64
DB :
236 AVGVINPAANTASRDQLVTRVTTHMAHALG-----FSVGPFEGARILLESINVRH 288

QY 65 -----KSEITITQSXYNGADV-----CGGADNSTIEL 92
DB :
289 KDPDVPVINSSTAVAKREYQGCDTLEYLEIEDQGAGSAGSHIKM 334

RESULT 14
F70675
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70675
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.V.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70675
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-582 <COL>
A:Cross-references: GB:T82098; GB:AL123456; NID:G3261664; PIDN:CAB05045.1; PID:e280737;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 13.1%; Score 99.5; DB 2; Length 582;
Best Local Similarity 23.4%; Pred. No. 0.35;
Matches 46; Conservative 19; Mismatches 77; Indels 55; Gaps 7;

QY 8 AFAAIVVSG-----SALAGVDQLVTRVTTHMAHASPDS----- 43
DB :
116 AFAQLVMSWFGNLAPLIAVEGAQEQAADVAAMVGHSGASAAAEOLVFPQALQOL 175

QY 44 -TLSIYQYGSAHALAQSDARKSETTI--TOSGYGN-----GADVQQ 83
DB :
176 PNLGIGNIGNALGGTGDLNTGNGNIGTNLGSNGRGDANLGSNGNIGNSNVGGGVGN 235

QY 84 ---GADNSTIEL----TQGFERNATIDWNKAHSKITVGOYGGNAALVNQTASD--- 132
DB :
236 GNFGSGNRAGLPGSGNVGNGMLGNSLTGSGTGNSSNVGFGTNNNVGTGNAGSNIGA 295

QY 133 SSVMVREQVGFNNATAN 149
DB :
296 GNTGSSNWGFNGNGIGN 312

RESULT 15
S19916
leishmanalysin (EC 3.4.24.36) precursor - Leishmania mexicana
N/Alternate names: surface metalloproteinase glycoprotein gp63
C:Species: Leishmania mexicana
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S19916; A48564
R:Medina-Acosta, E.; Karsess, R.E.; Russell, D.
submitted to the EMBL Data Library, February 1992
A:Description: Structurally distinct genes for the surface protease (gp63) of Leishmani
A:Reference number: S19916

Tue Aug 3 10:54:34 2004

A:Accession: S19916
A:Molecule type: mRNA
A:Residues: 1-646 <MED>
A:Cross-references: EMBL:X64394; NID:99559; PIDN:CAA45733.1; PID:99560
R:Medina-Acosta, E.; Kress, R.E.; Russell, D.G.
Mol. Biochem. Parasitol. 57, 31-45, 1993
A:Title: Structurally distinct genes for the surface protease of Leishmania mexicana ara
A:Reference number: A48564; MUID:93149206; PMID:8426614
A:Accession: A48564
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-176, Q' 178-646 <ME2>
A:Cross-references: EMBL:X64394; NID:99559
A>Note: sequence extracted from NCBI backbone (NCBIP:123747)
C:Genetics:
A:Gene: gp63-C1
A:Map position: 700kb chromosomal band
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds between two hydropho
A>Note: the activated form can activate the proenzyme form
C:Superfamily: leishmanolysin
C:Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; 2
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-102/Domain: activation peptide #status predicted <ATP>
F:103-646/Product: leishmanolysin #status predicted <MAT>
F:604-620/Domain: transmembrane #status predicted <TM1>
F:48,466,270,336/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
F:86,297,399,409,433,445,466,501/Binding site: carboxylate (Asn) (covalent) #status pre
F:127-144,193-232,316-388,395-458,408-427,417-492,469-513,518-568,538-561/disulfide bond
F:266,270,336/Binding site: zinc, catalytic (His) (active) #status predicted
F:267/Active site: Glu #status predicted

Query Match 13.1%; Score 99.5; DB 1; Length 646;
Best Local Similarity 30.8%; Pred. No. 0.4;
Matches 40; Conservative 15; Mismatches 66; Indels 9; Gaps 4;

QY 8 AFAAIVVGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQSDARK-- 65
Db 241 AVGVINIPAANIASTRYDQLVTRVTHMAHAGVSGTF---FGAVGIVQEVPHLRKDF 296
QY 66 SETTITQSGVNGADVGGADN-STIELTQNGFRNNA--TIDQWAKNSDITVGYGGNN 122
Db 297 NVSVITTSSTVAKAREQYGCNSLEYLEIEDQGGAGSAGSHIKVRNAKDELMAPAASAGYY 356
QY 123 AALVNOTASD 132
Db 357 TALTMVAFQD 366

Search completed: August 2, 2004, 14:56:22
Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds

(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRFQVGGNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 658 | 86.9 | 151 | 1 CSGA_SALTY | P55225 salmonella |
| 2 | 484 | 63.9 | 151 | 1 CSGA_ECOLI | P38307 escherichia |
| 3 | 474.5 | 62.7 | 152 | 1 CSGA_ECO57 | Q30424 escherichia |
| 4 | 118.5 | 15.7 | 151 | 1 CSGB_ECOLI | P39828 escherichia |
| 5 | 111.5 | 14.7 | 151 | 1 CSGB_SALTY | Q827m3 salmonella |
| 6 | 111.5 | 14.7 | 151 | 1 CSGB_SALTY | P55226 salmonella |
| 7 | 101.5 | 13.4 | 590 | 1 GP63_LEIDO | P32223 leishmania |
| 8 | 101 | 13.3 | 599 | 1 GP63_LEICH | P15706 leishmania |
| 9 | 93.5 | 13.1 | 646 | 1 GP63_LEIME | P43150 leishmania |
| 10 | 99 | 13.1 | 1656 | 1 OMPB_RICUA | Q06653 r outer mem |
| 11 | 97.5 | 12.9 | 1034 | 1 ICEN_PANAN | Q47879 pantoaea ana |
| 12 | 96.5 | 12.7 | 602 | 1 GP63_LEIMA | P08148 leishmania |
| 13 | 95.5 | 12.6 | 1258 | 1 ICEN_ERWHE | P16239 erwinia her |
| 14 | 94.5 | 12.5 | 1322 | 1 ICEA_PANAN | P20469 pantoaea ana |
| 15 | 93.5 | 12.4 | 1567 | 1 ICEN_XANCT | P18127 xanthomonas |
| 16 | 93 | 12.3 | 310 | 1 H8LO_STRMU | Q8cwx3 streptococc |
| 17 | 91.5 | 12.1 | 1655 | 1 OMPB_RICCN | Q9kka3 r outer mem |
| 18 | 87 | 11.5 | 331 | 1 OMB1_NEIMB | P30687 neisseria m |
| 19 | 87 | 11.5 | 331 | 1 OMB3_NEIMB | P30689 neisseria m |
| 20 | 85.5 | 11.3 | 1210 | 1 ICEN_PSEPL | P09815 pseudomonas |
| 21 | 84.5 | 11.2 | 1196 | 1 ICEV_PSEBX | P33479 pseudomonas |
| 22 | 84 | 11.1 | 823 | 1 NSP1_YEAST | P14907 saccharomyc |
| 23 | 83.5 | 11.0 | 1148 | 1 ICEK_PSEBX | Q30611 pseudomonas |
| 24 | 83.5 | 11.0 | 1200 | 1 ICEN_PSEBX | P06620 pseudomonas |
| 25 | 82.5 | 10.9 | 504 | 1 NRFI_BRARE | Q90x44 brachydanio |
| 26 | 82 | 10.8 | 955 | 1 FRU_PROMO | Q8in81 drosophila |
| 27 | 82 | 10.8 | 1025 | 1 SLAP_CAUCR | P35828 caulobacter |
| 28 | 81.5 | 10.8 | 422 | 1 HEAD_BPAPS | Q9t1s4 bacterioph |
| 29 | 81.5 | 10.8 | 760 | 1 YBIL_ECOLI | P25780 escherichia |
| 30 | 81.5 | 10.8 | 1158 | 1 RL14_HUMAN | P48552 homo sapien |
| 31 | 81 | 10.7 | 856 | 1 AUE7_AZOVI | Q32f99 azotobacter |
| 32 | 80.5 | 10.6 | 720 | 1 GTAC_BREDI | Q915dc brevundimon |
| 33 | 80 | 10.6 | 503 | 1 NRFI_HUMAN | Q16656 homo sapien |

| | | | | | |
|----|------|------|------|--------------|--------------------|
| 34 | 80 | 10.6 | 503 | 1 NRFI_MOUSE | Q9wu00 mus musculu |
| 35 | 80 | 10.6 | 621 | 1 GP63_LEIGU | Q00689 leishmania |
| 36 | 79.5 | 10.5 | 311 | 1 7B4C_PSESP | P07662 pseudomonas |
| 37 | 79.5 | 10.5 | 1300 | 1 120K_RICRI | P14914 rickettsia |
| 38 | 79.5 | 10.5 | 1654 | 1 OMPB_RICRI | Q53047 r outer mem |
| 39 | 79 | 10.4 | 907 | 1 GLRI_MOUSE | P23818 mus musculu |
| 40 | 79 | 10.4 | 907 | 1 GLRI_RAT | P19490 rattus norv |
| 41 | 78 | 10.3 | 582 | 1 Y281_MYCPN | P75383 mycoplasma |
| 42 | 78 | 10.3 | 1643 | 1 OMPB_RICPR | Q53020 r outer mem |
| 43 | 77 | 10.2 | 331 | 1 OMB_NEIMB | P30690 neisseria m |
| 44 | 76.5 | 10.1 | 388 | 1 ARGJ_CORGL | Q59280 corynebacte |
| 45 | 76 | 10.0 | 303 | 1 PANE_HALNI | Q9hxf0 halobacteri |

ALIGNMENTS

RESULT 1
CSGA_SALTY
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).

RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=1264504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

"Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.".

J. Bacteriol. 185:2330-2337(2003).

[5]

SEQUENCE FROM N.A.

RN SPECIES=S.enteritidis; STRAIN=27655-3B;

RP MEDLINE=9614651; PubMed=8550497;

RC COLLINSON S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;

RX "Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae";

RT J. Bacteriol. 178:662-667(1996).

RT [6]

SEQUENCE OF 21-151 FROM N.A.

RN SPECIES=S.enteritidis; STRAIN=27655-3B;

RP MEDLINE=94013373; PubMed=8104955;

RC DORAN J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,

RX MUNO C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;

RA "DNA-based diagnostic tests for Salmonella species targeting agfA, the structural gene for thin, aggregative fimbriae.";

RT J. Clin. Microbiol. 31:2263-2273(1993).

RT [7]

SEQUENCE OF 21-33.

RN SPECIES=S.enteritidis; STRAIN=27655-3B;

RP MEDLINE=91310586; PubMed=1677357;

RC COLLINSON S.K., Emeedy L., Mueller K.-M., Trust T.J., Kay W.W.;

RA "Purification and characterization of thin, aggregative fimbriae from Salmonella enteritidis";

RT J. Bacteriol. 173:4773-4781(1991).

RT CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN

CC --!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----

EMBL; AJ002301; CA005317.1; -

DR EMBL; AE008749; AAL20074.1; -

DR EMBL; AL627269; CAD08268.1; -

DR EMBL; AE016840; AAC69399.1; -

DR EMBL; U43280; AAC43599.1; -

DR PIR; JC6039; JC6039.

DR StyGene; SG10608; csga.

KW Fimbria; Signal; Complete proteome.

FT SIGNAL 1 20

FT CHAIN 21 151

FT FTCH 134 151

FT CONFLICT 134 151

FT FTCH 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;

SQ

Query Match 86.9%; Score 658; DB 1; Length 151;

Best Local Similarity 90.1%; Pred. No. 4.9e-50;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIIVVSGSAGLVVDQLVTRVTTHAHASGPDSTLSIIYQGSANAALAQ 60
| | | | | : | | | | |
Db 1 MKLLKVAFAAIIVVSGSAGLVVPQWGNGNHGGNSGPDSTLSIIYQGSANAALAQ 60
| | | | | : | | | | |
QY 61 SDARKSETTIQTSGYGNGADVGGADNSTIELTQTGFNRNATIDOWNAKNSDIITVGQYG 120
| | | | | : | | | | |
Db 61 SDARKSETTIQTSGYGNGADVGGADNSTIELTQTGFNRNATIDOWNAKNSDIITVGQYG 120
| | | | | : | | | | |
QY 121 NNAALVNQTASDSSVMVRQVFGGNATANQY 151
| | | | | : | | | | |
Db 121 NNAALVNQTASDSSVMVRQVFGGNATANQY 151
| | | | | : | | | | |


```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR EMBL; D90788; S70788.
DR EcoGene; E311489; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> E (IN REF. 1). CRC64;
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 63.9%; Score 484; DB 1; Length 151;
Best Local Similarity 66.9%; Pred. No. 4.8e-35;
Matches 101; Conservative 18; Mismatches 32; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQVLTVRVTHMAHASGDPSTLSIYQGSNAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVLTVRVTHMAHASGDPSTLSIYQGSNAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGVGG 120
DB 61 TDARNSDLTITQHGCGGADYVGGSDSSIDLTRQGFNSATLDQNGKSHMTVKQFG 120

QY 121 NNAALVNQTSDDSSVNVVQVFGNNATANQY 151
DB 121 GNGRAAVDQTASNSTVNVTVQVFGNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 62.7%; Score 474.5; DB 1; Length 152;
Best Local Similarity 65.8%; Pred. No. 3.2e-34;
Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQVLTVRVTH-EMAHASGDPSTLSIYQGSNAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVLTVRVTH-EMAHASGDPSTLSIYQGSNAALAL 60
QY 60 QSDARKSETTITQSGYNGADYVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGVGG 119
DB 61 QADARNSDLTITQHGCGGADYVGGSDSSIDLTRQGFNSATLDQNGKSHMTVKQFG 120

QY 120 GNNALVNQTSDDSSVNVVQVFGNNATANQY 151
DB 121 GNGRAAVDQTASNSTVNVTVQVFGNNATAHQY 152

RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGA_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGA OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in Escherichia coli

```

```

RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=X12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RN Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=X12;
RC MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi I.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RN DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=0157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Rofai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Godbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RN Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN=0157:H7 / RIMD 050952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RN DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RX STRAIN=X12;
RC MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csqBA promoter in
RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS."
RN Mol. Microbiol. 13:1021-1032(1994).
RN [7]
RP COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90754; CAA62281.1; -.

```

```

DR EMBL; AE000205; AAC74125.1; -.
DR EMBL; D90741; BAA35831.1; -.
DR EMBL; AE005315; AAG55787.1; -.
DR EMBL; AP002554; BAB34842.1; -.
DR PIR; C90806; C90806.
DR PIR; G85665; G85665.
DR PIR; S70787; S70787.
DR EcoGene; EG12621; csqB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CDS 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;

Query Match 15.7%; Score 118.5; DB 1; Length 151;
Best Local Similarity 31.6%; Pred. No. 0.0015;
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy 35 MAHAGSDPSTLSIYQYGSAAALQSDARKSTTTTQSGYNGADVGGQADNSTIELTQ 94
Db 18 IAAAGYDLANSEYNF-----AVNELSKSFNQAAIIQAGTNNSAQLRQGSKLLAVVAQ 73
Qy 95 NGFENNATIDQNAKNSDIITVGOYGGNNAALVNOTASDSSVMVROVYGFNNATANOV 151
Db 74 EGSNRKAKIDQTDYNL-AVIDAGSANDASISQAGYGTAMIIKSGNKANITQY 129

RESULT 5
CSGB_SALTI STANDARD; PRT; 151 AA.
AC Q827M5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth I., Connor P.,
RA Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar typhi CT18."
RN Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TV2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodymani V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar typhi strains Ty2
RT and CT18."
RN J. Bacteriol. 185:2330-2337(2003).
RN [3]
RP -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90754; CAA62281.1; -.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to licenses@isb-sib.ch).

| | |
|----|--|
| CC | EMBL; AJ002301; CAA05316.1; -; |
| DR | EMBL; AEC08749; AAL20073.1; -; |
| DR | EMBL; U43280; AAC43598.1; -; |
| DR | PIR; JC6040; JC6040. |
| DR | StyGene; SG10609; cscB. |
| DR | Zimbrria; Signal; Complete proteome. |
| KW | SIGNAL 1 21 |
| FT | POTENTIAL. |
| CH | CHAIN 22 151 |
| SO | SEQUENCE 151 AA; 16182 MW; CQFC5430E6DD361D CRC64; |

| DR | EMBL | ORF280 | AA:43358..11 | 1 | 21 | POTENTIAL |
|----|-----------|---------|-------------------|------------------|--------|-----------------------|
| DR | PIR | JC6040 | JC6040 | 1 | 21 | MINOR CURLIN SUBUNIT. |
| DR | StyleGene | SG10609 | csGB | 1 | 21 | |
| KW | Rimbia | Signal | Complete proteome | 1 | 21 | |
| FT | SIGNAL | 1 | 21 | 1 | 21 | POTENTIAL |
| CH | CHAIN | 22 | 151 | 22 | 151 | MINOR CURLIN SUBUNIT. |
| SO | SEQUENCE | 151 AA | 16182 MW | COFC5430E5DD361D | CRC664 | |

```

      1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16182 MW; 30FC5430AE6DD361D CRC64;

      14.7%; Score 11.5; DB 1; Length 151;
Best Local Similarity 30.5%; Pred. No. 0.0058;
Matches 36; Conservative 17; Mismatches 46; Indels 19; Gaps 5;

QY 51 GSANAALALQSPARKSE-----TTTQSYGYNGADVQQ-GADNST-----IELTQ 94
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 GAGGTATATNYDILARSEYFNAYNEIYKSKSFNOAAITIGVGTDSARVROEGSKLSAVIS 73
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

| | Query Match | Score | DB 1 | Length |
|-----------------------|-------------|-------------------|-------|----------------|
| Best Local Similarity | 14.7%; | Score 111.5; | DB 1; | Length 151; |
| Best Local Similarity | 30.5%; | Pred. No. 0.0058; | | |
| Mismatches | 36; | Conservative | 17; | Mismatches 46; |
| | | | | Indels 19; |
| | | | | Gaps 5; |

[illegible]

| | | | |
|----------|--------------|---|----------------|
| | 95 | N6FRNNATIDQWAKNSDIT-VGQGGGNNAALVNTASDSSVMVRQVGFGNNTANQY | 151 |
| | : | : : : | : |
| | : | : : | : |
| | : | : : | : |
| D8 | 74 | E6GNRAKVDQ--AGNYNPAYIEQTGNANDASISAYSAGNSAAIIKGSGKNANTQY | 129 |
| | | | |
| RESULT 7 | | | |
| | GP63 LEIDO | | |
| ID | GP63 LEIDO | STANDARD; | PRT; 590 AA. |
| AC | B23223; | | |
| DM | G23223V;1993 | (not) | 20 (corrected) |

| | |
|------------|---|
| RESULT 7 | |
| GP63_LEIDO | |
| ID | GP63_LEIDO |
| STANDARD; | PRT; 590 AA. |
| P23223; | |
| AC | |
| DT | 01-NOV-1991 (Rel. 20, Created) |
| DT | 01-NOV-1991 (Rel. 20, Last sequence update) |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) |
| DE | Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease) |
| DE | (Major surface glycoprotein) (GP63 protein) (Promastigote surface |

AC P2323;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 OS Leishmania donovani.
 GN GP63.
 OS

DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
DE GP63.
GN
GN
OC Leishmania donovani.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.

OS
Leishmania donovani.
OC
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC
NCBI_TaxID=5661;
OX
[1]
RN
SEQUENCE FROM N. A.
RP
STRAIN=LV9;
RC
MEDLINE=92107220; PubMed=1762629;
RX
Webb J.R., Button L.L., McMaster R.W.;
RT
Heterogeneity of the genes encoding the major surface glycoprotein

RP SEQUENCE FROM N.A.
STRAINE-IV9;
MEDLINE-92107220; PubMed-1762629;
Webb J.R., Burton L.L., McWhorter R.W.;
"Heterogeneity of the genes encoding the major surface glycoprotein
of Leishmania donovani";
Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
in the mammalian host.
CC

"Heterogeneity of the genes encoding the major surface glycoprotein of *Leishmania donovani*."; 48:173-184 (1991).
 Mol. Biochem. Parasitol. 48:173-184 (1991).
 -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at p1 and p1', and basic residues at p2 and p3'. A model nonapeptide is cleaved at Ala-Tyr|-Leu-Lys-Lys.
 -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

-!- FUNCTION: has an integral role during the maturation of macrophages in the mammalian host.

-!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr|-Leu-Iys-Lys-. CC

-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity). CC

-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor. CC

-!- SIMILARITY: Belongs to peptidase family M8. CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC Cleaved at Ala 1 zinc-finger-nucleic acids. CC
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity). CC
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor. CC
 CC -!- SIMILARITY: Belongs to peptidase family M8. CC
 CC ----- CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation CC
 CC the European Bioinformatics Institute. There are no restrictions on its CC
 CC use by non-profit institutions as long as its content is in no way CC
 CC modified and this statement is not removed. Usage by and for commercial CC

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M600148; AAA29244.1; -.
 DR USSB; P08148; 11M1.
 DP

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M60048; AAA29244.1; -
 HSSP; P08148; 1LML.

```

-----
CC
DR   EMBL; M60048; AAA29244.1; -
DR   HSSP; P08148; 1LML.
-----

```

```

DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase M8.
DR Pfam; PF01457; Peptidase M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 87 ACTIVATION PEPTIDE.
FT CHAIN 88 565 LEISHMANOLYSIN.
FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 251 221 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 252 222 BY SIMILARITY.
FT METAL 255 225 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 129 BY SIMILARITY.
FT DISULFID 178 217 BY SIMILARITY.
FT DISULFID 301 373 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC...) (POTENTIAL).
FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;

Query Match 13.4%; Score 101.5; DB 1; Length 590;
Best Local Similarity 28.8%; Pred. No. 0.2;
Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;

QY 8 AFAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQVGSANALALQSDAR-- 65
DB 226 AVGVINIPAANIAASYDQLVTRVVTTHMAHALG-----FSVVFPRDARILESISNVRHKDF 281
QY 66 -----SETTITQS--GYNGA-----DVGQADNSTLQTQGNFRNATIDOMNAK 109
DB 282 DVPVNSTAVAKAREQYCGTGLEVMEDQGGAGSAGSHIKM-----RNAQ-DELWAP 334
QY 110 NSDITVQVGGNNAALVNQTA---SDSSVMVRQVGFQNNNA 146
DB 335 ASD--AGVYSALTMAIFQDLGFGYQADFS--KAEEMPWGRNA 371

RESULT 8
GP63-LEICH STANDARD; PRT; 599 AA.
AC F15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]_TaxID=44271;
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
RT Arg-Gly-Asp sequence.";
RL Mol. Biochem. Parasitol. 39:267-274(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
RA Wilson M.E.;

```

```

RT "Three distinct RNAs for the surface protease gp63 are differentially
RT expressed during development of Leishmania donovani chagasi
RT promastigotes to an infectious form.";
RL J. Biol. Chem. 267:1898-1895(1992).
CC -I- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -I- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Dys]-.
CC -I- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -I- SIMILARITY: Belongs to peptidase family M8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
CC EMBL; M80672; AAA29238.1; -.
CC EMBL; M82527; AAA29235.1; -.
CC PIR; A44951; A44951.
CC HSP; P08148; 1LML.
CC MEROPS; M08.001; -.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase M8; 1.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC...) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 13.3%; Score 101; DB 1; Length 599;
Best Local Similarity 30.2%; Pred. No. 0.23;
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;

QY 8 AFAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPDSTLSIYQVGSANALALQSDAR--- 64
DB 236 AVGVINIPAANIAASYDQLVTRVVTTHMAHALG-----FSVVFPRDARILESISNVRH 288
QY 65 -----KSETTITQSGVNGADV-----CQGDNSTIEL 92
DB 289 KDFDVPVNSTAVAKAREQYCGTGLEVMEDQGGAGSAGSHIKM 334

RESULT 9
GP63-LEIME STANDARD; PRT; 646 AA.
ID -GP63-LEIME
AC P43150;

```


QY 6 VAAFAIVVSGSALGVYDQLVTRVTHMAHASGPDSTLSIYQYSANAALALQSDARK 65
 Db 509 VLAAGAITLDGSA-----TITGIDNGGG-----GAALQSITLANDATK 547
 QY 66 SETTITQSG-----YNGADVGGADNSTIELTQGNFRNNATID----- 104
 Db 548 ---TLTGLGANTISANGGTFINQANGGTFIKLST--QNNIVVDCDLAATDGTGVVDASS 602
 QY 105 QWNAKNSDI--TVQYQGGNNAAL-----VNQTASDSSVMVQVGFNGNAT 147
 Db 603 LTNQAQLTISGTFIGTIGANNITLQGFNIGSSKTLNGGNVAINELVIGNGS 654

RESULT 11
 ICEN PANAN STANDARD; PRT; 1034 AA.
 ID ICEN PANAN STANDARD; PRT; 1034 AA.
 AC Q47879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inau.
 GN INAU.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OC NCBI_TaxID=553;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=KUIN-3;
 RA MEDLINE=94264407; PubMed=7764866;
 RX Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
 uredovora.";
 RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
 CC -|- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -|- SUBCELLULAR LOCATION: Outer membrane.
 CC -|- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
 CC PERIODICITY IS SUPERIMPOSED.
 CC MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -|- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D14992; BAA03636.1; --
 DR FIR; JC2143; JC2143.
 DR HSSP; P06620; IINA.
 DR InterPro; IPR000258; Ice nucleatn.
 DR Pfam; PF00818; Ice nucleation; 51.
 DR PRINTS; PR00327; ICNUCLEATN.
 DR PROSITE; PS00314; ICE NUCLEATION; 34.
 KW Ice nucleation; Repeat; Outer membrane.
 FT DOMAIN 162-993 OCTAPEPTIDE PERIODICITY.
 SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match 12.9%; Score 97.5; DB 1; Length 1034;
 Best Local Similarity 28.7%; Pred. No. 0.85;
 Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;

QY 12 IVVSGSALGVYDQLVTRVTHMAHASGPDSTLSIYQYSANAALALQSDARKSETTI 70
 Db 161 IATYGSTLSGTHQSQSLIAGYGSTETA---GDSSTL-IAGYGSTGTA-----GSDSTL 208

QY 71 TQSGYNGADVGGADNSTIELTQGNFRNNATIDQWNAKNSDITVQYQYG-----GNNAAL 125
 Db 209 V-AGYGSTQTAGESSQMA-----GYGSTQT-----GMRGSDLTAG-YGSTGTAGDSSSL 256
 QY 126 V-----NQTASDSSVMVQVGFNGNATANQ 150
 Db 257 IAGYGSTQTAGESSLT--AGYGSTQTATQK 284

RESULT 12
 GP63 LEIMA STANDARD; PRT; 602 AA.
 ID GP63 LEIMA STANDARD; PRT; 602 AA.
 AC P08148; P15906;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 10-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5664;
 RN [1]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Button L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of leishmania.";
 RL J. Exp. Med. 167:724-729(1988).
 RN [2]
 RN REVISIONS.
 RA Button L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589(1990).
 RN [3]
 RN GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
 RA Homans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 RT the Leishmania major promastigote surface protease.";
 RL J. Biol. Chem. 265:16355-16364(1990).
 RN [4]
 RX X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 RT leishmanolysin, the major surface metalloproteinase from Leishmania
 RT major.";
 RL Proteins 22:58-66(1995).
 RN [5]
 RX X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 RT leishmanolysin.";
 RL Structure 6:1035-1046(1998).
 CC -|- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -|- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-
 CC -|- COFACTOR: Binds 1 zinc ion per subunit.
 CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -|- FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -|- SIMILARITY: Belongs to peptidase family M8.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Y00547; CAA58673.1; -.
 PIR: P0221; P0222.
 PDB: 1MLL; 17-SEP-97.
 MEROPS: M08.001; -.
 InterPro: IPR006025; Pept_M_Zn_BS.
 InterPro: IPR001577; Peptidase_M8.
 Pfam: PF01457; Peptidase_M8; 1.
 PROSITE: PS00782; LSHMANOLYSIN.
 PRINTS: PS00142; ZINC_PROTEASE; 1.
 PROSITE: PS00142; ZINC_PROTEASE; 1.
 Hydrolase; Metalloproteinase; Glycoprotein; Metal-binding; Zinc;
 Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.

POTENTIAL.
ACTIVATION PEPTIDE.
LEISHMANOLYSIN.
REMOVED IN MATURE FORM.
ZINC (CATALYTIC).

ZINC (CATALYTIC).
ZINC (CATALYTIC).

N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
GPI-anchor amidated asparagine.

| | | | |
|----|----------|---------|-----|
| FT | TURN | 328 | 332 |
| FT | STRAND | 335 | 339 |
| FT | TURN | 337 | 339 |
| FT | TURN | 341 | 342 |
| FT | STRAND | 343 | 343 |
| FT | TURN | 344 | 345 |
| FT | STRAND | 353 | 353 |
| FT | HELIX | 356 | 364 |
| FT | TURN | 365 | 366 |
| FT | STRAND | 369 | 370 |
| FT | HELIX | 372 | 374 |
| FT | TURN | 380 | 383 |
| FT | HELIX | 386 | 390 |
| FT | STRAND | 394 | 395 |
| FT | TURN | 396 | 397 |
| FT | STRAND | 398 | 399 |
| FT | TURN | 402 | 404 |
| FT | STRAND | 414 | 414 |
| FT | TURN | 417 | 418 |
| FT | STRAND | 421 | 425 |
| FT | STRAND | 428 | 429 |
| FT | HELIX | 435 | 437 |
| FT | TURN | 443 | 444 |
| FT | STRAND | 445 | 446 |
| FT | TURN | 450 | 454 |
| FT | STRAND | 458 | 465 |
| FT | TURN | 466 | 467 |
| FT | HELIX | 470 | 472 |
| FT | TURN | 475 | 477 |
| FT | HELIX | 478 | 480 |
| FT | TURN | 485 | 486 |
| FT | STRAND | 487 | 494 |
| FT | STRAND | 496 | 496 |
| FT | TURN | 506 | 516 |
| FT | STRAND | 517 | 520 |
| FT | TURN | 521 | 525 |
| FT | TURN | 527 | 528 |
| FT | STRAND | 533 | 534 |
| FT | TURN | 537 | 538 |
| FT | STRAND | 540 | 542 |
| FT | HELIX | 543 | 545 |
| FT | TURN | 546 | 546 |
| FT | STRAND | 550 | 550 |
| FT | TURN | 552 | 553 |
| FT | STRAND | 555 | 557 |
| FT | HELIX | 561 | 565 |
| FT | TURN | 566 | 567 |
| FT | HELIX | 569 | 572 |
| FT | TURN | 573 | 573 |
| FT | SEQUENCE | 602 AA; | 633 |

Query Match 12.7%; Score 96.5; DB 1; Length 602;
Best Local Similarity 59.5%; Pred. No. 0.56;
Matches 22; Conservative 3; Mismatches 9; Indels

QY 8 AFAAIIVSGSALAGVVDQLVTRVVTHEMAHA--SGP 41
 | | : : : : : : : : : : : : : :
DB 239 AVGVNIPAAIASRYDQLVTRVVTHEMALGESGP 275

| | |
|------------|---|
| RESULT 13 | |
| ID | ICEN_ERWHE |
| ICEN_ERWHE | STANDARD; PRT; 1258 AA. |
| AC | PI6239; |
| DT | 01-APR-1990 (Rel. 14, Created) |
| DT | 01-APR-1990 (Rel. 14, last sequence update) |
| DT | 16-OCT-2001 (Rel. 40, last annotation update) |
| DE | Ice nucleation protein. |
| DE | Ice nucleation protein. |
| GN | ICEE. |
| GN | GN |
| OS | Erwinia herbicola. |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; |
| OC | Enterobacteriaceae; Pantoea. |
| OX | NCBI TaxID=549; |


```
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52970; CAA37140.1; -.
DR HSP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleatn; 81.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 57.
KW Ice_nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D9595ECAD63 CRC64;

Query Match 12.4%; Score 93.5; DB 1; Length 1567;
Best Local Similarity 29.9%; Pred. No. 3;
Matches 43; Conservative 26; Mismatches 30; Indels 45; Gaps 12;

QY 39 SGPDSTLSIYQVGSANAALQSDARKSETTITQSGYGN-----GADV-----GQG 84
Db 1060 AGADSTL- IAGYGSTQTA-----GSDSSLT-AGYGTQTARQGS DITAGYGTGTAG 1109

QY 85 ADNSTIE---LTQN-GPRNNATI---DOWNAKNSDITVGOYG-----GNNALV----- 126
Db 1110 ADSSLIAGYGTQTAGYDSNLTAGYGTQTAREDSLSITAG-YGSTSTAGHDSLSIAGYGS 1168

QY 127 NQTASDSSVMVRQVGFNNATANQ 150
Db 1169 TQTAGYNSILT--TGYGSTQTAQE 1190
```

Search completed: August 2, 2004, 14:49:29
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_arChaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|--------|--------------------|
| 1 | 649 | 85.7 | 152 | Q33802 | Q33802 salmonella |
| 2 | 563.5 | 74.4 | 150 | Q7X243 | Q7X243 citrobacter |
| 3 | 524 | 69.2 | 149 | Q7X240 | Q7X240 citrobacter |
| 4 | 471.5 | 62.3 | 152 | Q8CW63 | Q8CW63 escherichia |
| 5 | 454.5 | 60.0 | 150 | Q7X237 | Q7X237 enterobacte |
| 6 | 340 | 44.9 | 76 | Q54069 | Q54069 salmonella |
| 7 | 133 | 17.6 | 502 | Q8EIH4 | Q8EIH4 shewanella |
| 8 | 120 | 15.9 | 160 | Q8CW64 | Q8CW64 escherichia |
| 9 | 120 | 15.9 | 160 | Q83R07 | Q83R07 shigella fl |
| 10 | 118.5 | 15.7 | 151 | Q7UCZ1 | Q7UCZ1 shigella fl |
| 11 | 118.5 | 15.7 | 153 | Q89J16 | Q89J16 bradyrhizob |
| 12 | 116 | 15.3 | 171 | Q89J13 | Q89J13 bradyrhizob |
| 13 | 115 | 15.2 | 139 | Q8EIH3 | Q8EIH3 shewanella |
| 14 | 110 | 14.5 | 130 | Q89J14 | Q89J14 bradyrhizob |
| 15 | 108.5 | 14.3 | 151 | Q7X244 | Q7X244 citrobacter |
| 16 | 107.5 | 14.2 | 154 | Q89J15 | Q89J15 bradyrhizob |

| | | | | | | |
|----|-------|------|------|----|--------|--------------------|
| 17 | 105.5 | 13.9 | 91 | 2 | Q9S3J8 | Q9S3J8 escherichia |
| 18 | 102.5 | 13.5 | 152 | 2 | Q7X241 | Q7X241 citrobacter |
| 19 | 101.5 | 13.4 | 2087 | 5 | Q8MVW7 | Q8MVW7 naegleria g |
| 20 | 101 | 13.3 | 599 | 5 | Q25273 | Q25273 leishmania |
| 21 | 99.5 | 13.1 | 582 | 16 | P71868 | P71868 mycobacteri |
| 22 | 99.5 | 13.1 | 582 | 16 | Q7TW98 | Q7TW98 mycobacteri |
| 23 | 99.5 | 13.1 | 641 | 5 | Q8MN20 | Q8MN20 leishmania |
| 24 | 99.5 | 13.1 | 657 | 5 | Q8MN21 | Q8MN21 leishmania |
| 25 | 99 | 13.1 | 1613 | 2 | Q840U5 | Q840U5 rickettsia |
| 26 | 99 | 13.1 | 1618 | 2 | Q9KKB1 | Q9KKB1 rickettsia |
| 27 | 98.5 | 13.0 | 151 | 2 | Q7X238 | Q7X238 enterobacte |
| 28 | 98 | 12.9 | 145 | 16 | Q8U6N9 | Q8U6N9 agrobacteri |
| 29 | 98 | 12.9 | 157 | 16 | Q88HG0 | Q88HG0 pseudomonas |
| 30 | 97 | 12.8 | 1612 | 2 | Q840U6 | Q840U6 rickettsia |
| 31 | 97 | 12.8 | 1617 | 2 | Q7X5N9 | Q7X5N9 rickettsia |
| 32 | 96.5 | 12.7 | 348 | 13 | Q93397 | Q93397 cyprinus ca |
| 33 | 96.5 | 12.7 | 644 | 5 | Q43994 | Q43994 leishmania |
| 34 | 96.5 | 12.7 | 1306 | 2 | Q93N36 | Q93N36 pantoea ana |
| 35 | 96.5 | 12.7 | 1341 | 16 | Q8ED31 | Q8ED31 shewanella |
| 36 | 96.5 | 12.7 | 2016 | 5 | Q9BIT0 | Q9BIT0 plectreury |
| 37 | 96 | 12.7 | 598 | 5 | Q25275 | Q25275 leishmania |
| 38 | 96 | 12.7 | 641 | 5 | Q8MN22 | Q8MN22 leishmania |
| 39 | 96 | 12.7 | 641 | 5 | Q8MN48 | Q8MN48 leishmania |
| 40 | 96 | 12.7 | 641 | 5 | Q8MN23 | Q8MN23 leishmania |
| 41 | 96 | 12.7 | 641 | 5 | Q8MN24 | Q8MN24 leishmania |
| 42 | 96 | 12.7 | 641 | 5 | Q8MN25 | Q8MN25 leishmania |
| 43 | 95.5 | 12.6 | 179 | 2 | Q33801 | Q33801 salmonella |
| 44 | 95.5 | 12.6 | 617 | 5 | Q94593 | Q94593 leishmania |
| 45 | 95.5 | 12.6 | 1613 | 2 | Q9KKB2 | Q9KKB2 israeli tic |

ALIGNMENTS

RESULT 1

Q33802 : PRELIMINARY; PRT; 152 AA.
ID O33802 :
AC O33802 :
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Agfa protein (fragment).
GN AGFA.
GN Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CAA04151.1; -.
FT NON_TER 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 85.7%; Score 649; DB 2; Length 152;

Best Local Similarity 88.7%; Pred. No. 1.1e-44;

Matches 134; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | MKLLKVAFAAIVVSGSALAGVVDQLVTRVWVTHEMAHAGSPDSTLSIYGYGSANAALAQ | 60 |
| DB | 1 | MKLLKVAFAAIVVSGSALAGVVDQLVTRVWVTHEMAHAGSPDSTLSIYGYGSANAALAQ | 60 |
| QY | 61 | SDARKSETTITGSGYNGADVQGGADNNTIELTQNGFRNNATIDQNAKNSDITVQYGG | 120 |
| DB | 61 | SDARKSETTITGSGYNGADVQGGADNNTIELTQNGFRNNATIDQNAKNSDITVQYGG | 120 |
| QY | 121 | NNAALVNTQATSDSSVMVRQVGFNNATANQY | 151 |

```

Db      121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 74.4%; Score 563.5; DB 2; Length 150;
Best Local Similarity 74.2%; Pred. No. 8e-38;
Matches 118; Conservative 11; Mismatches 13; Indels 17; Gaps 2;

QY 1 MKLLKVAAPAAIIVSGSALAGYVDQLVTRVWTHM-----AHASGPDSTLSIYQVGS 52
|||||
DB 1 MKLLQVAAPAAIIVSGSALAGSVQP-----WGGGGGGGGSSSGPESTLSIYQSGV 51
|||||
QY 53 ANAALALQSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSD 112
|||||
DB 52 NNAALALQSDARKSDTTIHQNGFGNGADVGQGSDDSTIDLTQRFNGSATLDQWNGKDSMTVTKQFG 111
|||||

QY 113 ITVGQVGGNNAALVNQTASDSSVMVRQVFGNNATANQY 151
|||||
DB 112 ITVSQVGGHNAALVNQTASDSSVLVHQVFGNNATANQY 150
|||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.2%; Score 524; DB 2; Length 149;
Best Local Similarity 72.8%; Pred. No. 1.2e-34;
Matches 110; Conservative 14; Mismatches 25; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGYVDQLVTRVWTHMNAHASGPDSTLSIYQVGSANAALAQ 60
|||||

```

```

Db      1 MKLLKVAAPAAIIVSGSALAGYVVPQWGN--HHGGGSNYGPDSSLSIYQVGSNNSANALQ 58
61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYQG 120
|||||
DB 59 SDARKSDVTITQHGRNGAVVGQGGDDSTLSLQGTGFSATLDQWNAKNSDITVTKQFG 118
|||||
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
|||||
DB 119 NGALVNQTASDSNVLIQVFGNNATANQH 149
|||||

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 62.3%; Score 471.5; DB 16; Length 152;
Best Local Similarity 65.8%; Pred. No. 2e-30;
Matches 100; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGYVDQLVTRVWTH-EMAHASGPDSTLSIYQVGSANAALAL 59
|||||
DB 1 MKLLKVAAPAAIIVSGSALAGYVPPVGGGGHGGGNGNSSEINLIYQGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYQG 119
|||||
DB 61 QADARNSDLTITQHGCGNGADVGQGSDDSSIDLQRFNGSATLDQWNGKDSMTVTKQFG 120
|||||
QY 120 GNAALVNQTASDSSVMVRQVFGNNATANQY 151
|||||
DB 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 152
|||||

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 508BB2D872DF15F3 CRC64;

Query Match 60.0%; Score 454.5; DB 2; Length 150;
Best Local Similarity 62.1%; Pred. No. 4.6e-29;
Matches 95; Conservative 24; Mismatches 29; Indels 5; Gaps 2;

QY 1 MKLKVAAFAAIVVSGSALAGVYDQVLTFRVVVTHEMAHA--SGPDSTLSIYQYGSANAALA 58
DB 1 MKFKVAAALAAIVVSGSMAGVNIQ---GGWGHGGHGGYGGPNTLNIYQNGGNSALA 57

QY 59 LOSDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATIDQWNAKNSDITVQY 118
DB 58 LQTDARNVNLISOTGGGNGADVGCGDSSINITQNGFGNSATLDQWNSKDSVMNVSY 117

QY 119 GGNNAALVNQNTASDSSVMVROVGGNNATANQY 151
DB 118 GGLNGALVDQFASNSTVNTVQIGFNGHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DE 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SEF17 fimbrin (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 44.9%; Score 340; DB 2; Length 76;
Best Local Similarity 97.1%; Pred. No. 3.1e-20;
Matches 67; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGCGADNSTIETQNG 96
DB 8 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGCGADNSTIETQNG 67

QY 97 FRNNATIDQ 105
DB 68 FRNNATIDQ 76

RESULT 7
Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SO0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.

```

```

OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22397686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Isapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015332; AAN53941.1; -.
DR TIGR; SO0865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 17.6%; Score 133; DB 16; Length 502;
Best Local Similarity 29.6%; Pred. No. 0.013;
Matches 37; Conservative 21; Mismatches 59; Indels 8; Gaps 3;

QY 32 THEMAHAGPDPSTLSIYQYGSANA---ALALQSDARKSETTITQSGYNGADVGCGADNS 88
DB 258 TAYLSMTTGGDNTVDITQDSDNTVGVDSLIADIQGDNDNITIKQKDSNGAEFQVWGDSN 317

QY 89 TIELTQNGFRNATIDQWNAKNSDITVQYGGNNAALVNQNTASDSSVMVROVQFGN---- 144
DB 318 DVDLQKQGDANFATFGAYGTDN--DFDLSSKGDNNELVAFATGEDNSIEISQEGDANFAYV 376

QY 145 NATAN 149
DB 377 DATGN 381

RESULT 8
Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;
Best Local Similarity 28.7%; Pred. No. 0.036;
Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;

QY 22 VYDQ-----LTVRVVTHEMAHAGPDPSTLSIYQYGSANAALQSDARKSET 69
DB 1 MYDQVQGDNMKNKLLFMVLTILGAPGIAAAAGYDLANSEYNF----AVNELSKSFFNQAA 56

```

QY 69 TITQSGYNGADVGQAGNDSITLTONGFRNNATIDQNNAKNSDITVQYGGNNAALVNO 128
 Db 57 IICQAGTNNNSAQIRQGGSKLLTVVAGGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQ 115
 QY 129 TASDSSVMVRQVGFNNATANQY 151
 Db 116 GAYGNTAMIIQKSGNKNITQY 138

RESULT 9

Q83RU7 PRELIMINARY; PRT; 160 AA.
 AC Q83RU7;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Minor curlin subunit precursor, similar to CsgA.
 GN CSGB OR SF1035.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157."
 RL Nucleic Acids Res. 30:4432-4441(2002).
 DR EMBL; A5015131; AAN42658.1; -.
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;
 Best Local Similarity 28.7%; Pred. No. 0.036;
 Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;
 QY 22 VYDQ-----LVTVVTHMAHAGSPDSTLSIYQYGSANAALAQSDARKSET 68
 Db 1 MYDQVGDNMKNKLLFMMLTILGAGIAAAGYDLANSEYNF----AVNELSKSFSNQAA 56
 QY 69 TITQSGYNGADVGQAGNDSITLTONGFRNNATIDQNNAKNSDITVQYGGNNAALVNO 128
 Db 57 IICQAGTNNNSAQIRQGGSKLLTVVAGGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQ 115
 QY 129 TASDSSVMVRQVGFNNATANQY 151
 Db 116 GAYGNTAMIIQKSGNKNITQY 138

RESULT 10

Q7UCZ1 PRELIMINARY; PRT; 151 AA.
 AC Q7UCZ1;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Minor curlin subunit.
 GN CSGB OR S1108.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T."
 RL Infect. Immun. 71:2775-2786(2003).
 DR EMBL; AE016981; AAP16542.1; -.
 SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 151;
 Best Local Similarity 31.6%; Pred. No. 0.045;
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

QY 35 MAHAGSPDSTLSIYQYGSANAALAQSDARKSETTITQSGYNGADVGQAGNDSITLTONG 94
 Db 18 IAAAGYDLANSEYNF----AVNELSKSFSNQAAIICQAGTNNNSAQIRQGGSKLLAVVQA 73
 QY 95 NGFRNNATIDQNNAKNSDITVQYGGNNAALVNQTSDSVMVRQVGFNNATANQY 151
 Db 74 EGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQAGYGNNTAMIIQKSGNKNITQY 129

RESULT 11

Q89JI6 PRELIMINARY; PRT; 153 AA.
 AC Q89JI6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE CsgB protein.
 GN CSGB OR BLIS297.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiuni T.,
 RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
 RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
 RA Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 RT Bradyrhizobium japonicum USDA110."
 RL DNA Res. 9:189-197(2002).
 DR EMBL; AP005954; BAC50562.1; -.
 KW Complete proteome.
 SQ SEQUENCE 153 AA; 15991 MW; 4CE71DEAC375145B CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 153;
 Best Local Similarity 27.0%; Pred. No. 0.046;
 Matches 41; Conservative 32; Mismatches 56; Indels 23; Gaps 6;

QY 2 KLIKVAA-FAAIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTL-SIYQYGSANAAL 59
 Db 10 RVLVAALLAAGCAATQASAGSIQSVT-----NPNVSIETIVQFGNDVQPVTI 58
 QY 60 QSDARKSETTITQSGYNGADVGQAGNDSITLTONGFRNNATIDQNNAKNSDITVQYGG 119
 Db 59 EENSrvNIARVQIG-----GSGTVDATI--IQNGTRNYANVIMGG-TTNAAGVQSG 108
 QY 120 GNNALVNQTSDSVMVRQVGFNNATANQY 151
 Db 109 LNSTADITQIGNSTALLIQIDMNSGAVRQF 140

RESULT 12

Q89JI3 PRELIMINARY; PRT; 171 AA.
 AC Q89JI3;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)


```

RC STRAIN=Fec2;
RA Zcgaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD58671.1; -. BD00AF57E1400704 CRC64;
SQ SEQUENCE 151 AA; 16158 MW;
Query Match 14.3%; Score 108.5; DB 2; Length 151;
Best Local Similarity 27.4%; Pred. No. 0.29;
Matches 32; Conservative 20; Mismatches 48; Indels 17; Gaps 3;
QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVGGADNSTIELTQNGFR----- 98
DB 14 GAPGIASATSYDLAHSEYNFAYNELSKGSFNQAIIIGQVGTNNSAKMRQEGSKLLSVVSQ 73
QY 99 ----NNATIDOWNAKNSDITVGQYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151
DB 74 EGGSNPAKYDQSGAYNF-AYIAQSGHSNDASISQSNYGNNTAMIIQKSGSNKANITQY 129

```

Search completed: August 2, 2004, 14:54:38
Job time : 30.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780
Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04.*

1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|--------------------|
| 1 | 780 | 100.0 | 151 | 3 | AAB36349 | AAB36349 Agfa::PT3 |
| 2 | 691 | 88.6 | 151 | 2 | AAR74625 | Aar74625 Agfa sequ |
| 3 | 691 | 88.6 | 151 | 3 | AAB36341 | AAB36341 Salmonell |
| 4 | 686 | 87.9 | 151 | 2 | AAW23570 | AAW23570 Salmonell |
| 5 | 675 | 86.5 | 151 | 3 | AAB36353 | AAB36353 Agfa::PT3 |
| 6 | 622 | 78.7 | 151 | 3 | AAB36350 | AAB36350 Agfa::PT3 |
| 7 | 618 | 78.2 | 151 | 3 | AAB36348 | AAB36348 Agfa::PT3 |
| 8 | 613 | 78.6 | 151 | 3 | AAB36346 | AAB36346 Agfa::PT3 |
| 9 | 611 | 78.3 | 151 | 3 | AAB36347 | AAB36347 Agfa::PT3 |
| 10 | 608 | 77.9 | 151 | 3 | AAB36352 | AAB36352 Agfa::PT3 |
| 11 | 601 | 77.1 | 151 | 3 | AAB36354 | AAB36354 Agfa::PT3 |
| 12 | 600 | 76.9 | 151 | 3 | AAB36351 | AAB36351 Agfa::PT3 |
| 13 | 599 | 76.8 | 151 | 3 | AAB36355 | AAB36355 Agfa::PT3 |
| 14 | 523 | 67.1 | 151 | 3 | AAB36343 | AAB36343 Escherich |
| 15 | 518 | 66.4 | 151 | 7 | ABR82651 | ABR82651 E. coli C |
| 16 | 506 | 64.9 | 120 | 2 | AAR62761 | Aar62761 Agfa sequ |
| 17 | 506 | 64.9 | 120 | 2 | AAW23569 | AAW23569 Salmonell |
| 18 | 445 | 57.1 | 142 | 2 | AAR52664 | Aar52664 Fibronect |
| 19 | 373 | 47.8 | 122 | 2 | AAR52663 | Aar52663 FNB curli |
| 20 | 237 | 30.4 | 45 | 3 | AAB36316 | AAB36316 Salmonell |
| 21 | 132 | 16.9 | 22 | 3 | AAB36318 | AAB36318 Salmonell |
| 22 | 123 | 15.8 | 23 | 3 | AAB36321 | AAB36321 Salmonell |
| 23 | 123 | 15.8 | 23 | 3 | AAB36326 | AAB36326 Salmonell |
| 24 | 123 | 15.8 | 23 | 3 | AAB36338 | AAB36338 Salmonell |
| 25 | 115 | 14.7 | 22 | 3 | AAB36325 | AAB36325 Salmonell |

| | | | | | | |
|----|-------|------|-----|---|----------|--------------------|
| 26 | 115 | 14.7 | 22 | 3 | AAB36339 | AAB36339 Salmonell |
| 27 | 115 | 14.7 | 22 | 3 | AAB36320 | AAB36320 Salmonell |
| 28 | 111 | 14.2 | 22 | 3 | AAB36322 | AAB36322 Salmonell |
| 29 | 111 | 14.2 | 22 | 3 | AAB36327 | AAB36327 Salmonell |
| 30 | 111 | 14.2 | 22 | 3 | AAB36337 | AAB36337 Salmonell |
| 31 | 109 | 14.0 | 22 | 7 | ABR82644 | ABR82644 E. coli c |
| 32 | 107.5 | 13.8 | 151 | 3 | AAB36344 | AAB36344 Escherich |
| 33 | 106.5 | 13.7 | 151 | 3 | AAB36342 | AAB36342 Salmonell |
| 34 | 102 | 13.1 | 26 | 7 | ABR82649 | ABR82649 E. coli V |
| 35 | 101 | 12.9 | 262 | 4 | ABR86342 | ABR86342 Drosophil |
| 36 | 98 | 12.6 | 26 | 7 | ABR82645 | ABR82645 E. coli c |
| 37 | 96 | 12.3 | 19 | 3 | AAB36323 | AAB36323 Salmonell |
| 38 | 96 | 12.3 | 19 | 3 | AAB36336 | AAB36336 Salmonell |
| 39 | 96 | 12.3 | 19 | 3 | AAB36328 | AAB36328 Salmonell |
| 40 | 95 | 12.2 | 24 | 7 | ABR82647 | ABR82647 E. coli c |
| 41 | 93.5 | 12.0 | 287 | 4 | ABR86343 | ABR86343 Drosophil |
| 42 | 93 | 11.9 | 974 | 6 | ABU17075 | ABU17075 Protein e |
| 43 | 92 | 11.8 | 23 | 3 | AAB36331 | AAB36331 Escherich |
| 44 | 92 | 11.8 | 975 | 6 | ADA33477 | ADA33477 Aginetoba |
| 45 | 91.5 | 11.7 | 580 | 5 | AAG66008 | AAG66008 F. necrop |

ALIGNMENTS

RESULT 1
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX AAB36349;
AC AAB36349;
XX
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64625.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 780; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.7e-67; Indels 0; Gaps 0;
 Matches 151; Conservative 0; Mismatches 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYDQLVTRVVTHEMAHALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120

QY 121 NNAALVNQTSASSVMVRQVGFNNATANOY 151
 DB 121 NNAALVNQTSASSVMVRQVGFNNATANOY 151

RESULT 2
 AAR74625
 ID AAR74625 standard; protein; 151 AA.

AC AAR74625;
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)

XX Agfa sequence.

XX Salmonella; Agfa; vaccine.

XX Salmonella.

XX WO9425598-A2.

XX 10-NOV-1994.

XX 26-APR-1994; 94WO-IB000207.

XX 26-APR-1993; 93US-00054452.

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;

XX WPI; 1994-358275/44.

DR N-PSDB; AAQ87467.

XX Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.

PS Disclosure; Fig 7B; 95pp; English.

XX The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 151 AA;

Query Match 88.6%; Score 691; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYDQLVTRVVTHEMAHALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQGSANALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120

DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120

QY 121 NNAALVNQTSASSVMVRQVGFNNATANOY 151

DB 121 NNAALVNQTSASSVMVRQVGFNNATANOY 151

RESULT 3

AAB36341

ID AAB36341 standard; protein; 151 AA.

XX AAB36341;

XX 26-FEB-2001 (first entry)

XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

XX vaccine; immune response; immunogen.

XX Salmonella enteritidis.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collinson SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64617.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 88.6%; Score 691; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYLQVTRVVTHEMAHALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQ 60

Qy 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 Db 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

Qy 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
 Db 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 4

AAW23570
 ID AAW23570 standard; protein; 151 AA.

XX AC AAW23570;

XX 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)

XX Salmonella enteritidis 27655-3b agfa.

XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.

XX Salmonella enteritidis.

XX Key Location/Qualifiers
 FH Misc-difference 123
 FT /note= "Encoded by GCC"

XX US5635617-A.

XX 03-JUN-1997.

XX 26-APR-1994; 94US-00233788.

XX 26-APR-1993; 93US-00054452.

XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX Collinson SK, Kay WW, Doran JL;

XX WPI; 1997-309886/28.

XX N-PSDB; AAT74142.

XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.

XX Example 2; Fig 7; 85pp; English.

XX The present sequence represents agfa encoded by the full agfa gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)

XX Sequence 151 AA;

Query Match 87.9%; Score 686; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.1e-58;
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDYLQVTRVVTHEMAHALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60

Qy 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 Db 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

Qy 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151
 Db 121 NNPALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 5

AA36353
 ID AA36353 standard; protein; 151 AA.

XX AC AA36353;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

XX Escherichia coli.

XX Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collinson SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64629.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 86.5%; Score 675; DB 3; Length 151;
 Best Local Similarity 81.9%; Pred. No. 2.4e-57;
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPD-----YDQ 45
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPDSTLSIYQYGSANAALYDQ 60
 QY 46 LVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGOGADNSTIELTQNGFRNNATIDQ 105
 DB 61 LVTRVVTHEMAHA-----GYNGADVGOGADNSTIELTQNGFRNNATIDQ 105
 QY 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 6

AAB36350
 ID AAB36350 standard; protein; 151 AA.
 AC
 XX AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN W0200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PF which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT

PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/FAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC copy of that gene; and (4) eliciting an immune response in an animal, Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 79.7%; Score 622; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 3.3e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPD----- 42
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGHNGGNSGGPDSTLSIYQYGSANAALQ 60
 QY 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGOGADNSTIELTQNGF 97
 DB 61 SDARKYDQLVTRVVTHEMAHA-----GGADNSTIELTQNGF 97
 QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7

AAB36348
 ID AAB36348 standard; protein; 151 AA.
 AC
 XX AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.
 XX
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN W0200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX

```
PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64624.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of a recombinant gene into the chromosome of the
CC homologous species; (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 79.2%; Score 618; DB 3; Length 151;
XX Best Local Similarity 76.6%; Pred. No. 8e-52;
XX Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAH--- 57
XX 1 MKLLKVAFAAIVVSGSALAGV-----YDQLVTRVVTHEMAHAGS 40
XX
XX 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
XX 41 PDSTLSIYQGSANAALAQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
XX
XX 101 ATTDOWNAKNSDITVQYGGNNAALVNOTASDSSVWVQVGFNNATANQY 151
XX 101 ATTDOWNAKNSDITVQYGGNNAALVNOTASDSSVWVQVGFNNATANQY 151
XX
XX RESULT 8
XX ID AAB36346
XX
XX AAB36346 standard; protein; 151 AA.
XX
XX AAB36346;
XX
XX 26-FEB-2001 (first entry)
XX
XX AgfA::PT3#1 amino acid sequence SEQ ID NO:12.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.6%; Score 613; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 2.4e-51;
XX Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALAQ 60
XX
XX 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATTDOWNAKNSDITVQYGG 120
XX 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATTDOWNAKNSDITVQYGG 120
XX
XX 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
XX 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
XX
XX RESULT 9
```

AAB36347
 ID AAB36347 standard; protein; 151 AA.
 XX
 AC AAB36347;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEFI7/TAFF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

Query Match 78.3%; Score 611; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3.8e-51;
 Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVTTHENAHLQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVTTHENAHLQ 60

QY 61 SDARKSETTITQSGYNGADVQGCADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGCADNSTIELTQNGFRNNATIDQWNAKNSDITVQYDQ 120
 QY 121 NNAALVNQTSADSSVMVRQVGFQGNATANQY 151
 DB 121 LVTRVVTTHENAHSVMVRQVGFQGNATANQY 151

RESULT 10
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 XX
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEFI7/TAFF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.9%; Score 608; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 7.4e-51;
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQVGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVVT 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX
 AC AAB36354;
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200050102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64630.
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.1%; Score 601; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 3.5e-50;
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQVGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 XX
 AC AAB36351;
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#6 amino acid sequence SEQ ID NO:22.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEPI7/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA, and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 76.9%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4.4e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 AC AAB36355;
 XX
 XX 26-FEB-2001 (first entry)
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI WPI; 2000-672631/65.
 DR

DR N-PSDB; AAC64631.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 139; 139pp; English.
 XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEPI7/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA, and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 76.8%; Score 599; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDYDQLVTRVVTHEMAHALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 AC AAB36343;
 XX
 XX 26-FEB-2001 (first entry)
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Escherichia coli.
 OS
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX


```
PF 05-APR-2000; 2000WO-CA000356.
XX
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX N-PSDB; AAC64619.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant AgfA
XX PT protein useful for eliciting immune response in animal.
XX
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX CC segment of the gene has been replaced by a segment of a foreign DNA
XX CC sequence which encodes a foreign epitope or antigen. Also described are:
XX CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
XX CC assembly system of strains of Salmonella, Escherichia coli and
XX CC Enterobacteriaceae for the production of fimbriae comprising recombinant
XX CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX CC directing recombination of a recombinant gene into the chromosome of the
XX CC homologous species; (3) directing recombination of a recombinant gene
XX CC back into the chromosome of the homologous species, replacing the native
XX CC copy of that gene; and (4) eliciting an immune response in an animal,
XX CC comprising separating an amino acid polymer comprising a recombinant AgfA
XX CC protein containing a replacement segment or segments of foreign amino
XX CC acid sequence or sequences grown on a Salmonella, E. coli or
XX CC Enterobacteriaceae host cell, from the host cell and introducing the
XX CC polymer into the animal in conjunction with a carrier or diluent. (I) is
XX CC useful for the expression of recombinant AgfA protein which is useful for
XX CC eliciting an immune response in an animal. In a fimbrial presentation
XX CC system the heterologous antigens are presented in high numbers (up to
XX CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX CC immunogenicity and adhesion properties relevant for an efficient live
XX CC vaccine, the carrier fimbrial subunit proteins are usually strong
XX CC immunogens, which may be important for directing an immune response
XX CC against the inserted epitope, and hybrid fimbriae are easy and
XX CC inexpensive to purify in large amount. The present sequence is given in
XX CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 67.1%; Score 523; DB 3; Length 151;
XX Best Local Similarity 68.9%; Pred. No. 1.2e-42; Mismatches 27; Indels 0; Gaps 0;
XX Matches 104; Conservative 20;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNNGSSGPDYDQLVTRVVTHEMAHALQ 60
XX Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNNGSSGPDYDQLVTRVVTHEMAHALQ 60
XX
XX QY 61 SDARKSETTITQSGYGNAGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
XX Db 61 TDARNSDLTITQGGNGADVCGGSDSSIDLITQRFNGSATLDQNGKNSMTVKQFGG 120
XX
XX QY 121 NNAALVNQTASDSSVVMVQVGFNNATANQY 151
XX Db 121 GNGAAVDQTASNSSVNVTVQVGFNNATAHQY 151
XX
XX RESULT 15
XX ABR82651
XX ID ABR82651 standard; protein; 151 AA.
XX
XX AC ABR82651;
XX
XX XX
XX DT 04-DEC-2003 (first entry)
XX
XX DE E. coli CsgA subunit 15 kDa protein.
XX
XX
```

```
XX
XX Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
XX Escherichia coli.
XX
XX WO2003064446-A2.
XX
XX 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EF000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-646136/61.
XX
XX N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
XX PT useful in the manufacture of a medicament for the prevention and/or
XX PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
XX PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
XX CC mammalian plasma protein or of generating an immune response in a mammal
XX CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
XX CC antibody is useful for treating a bacterial infection in a human or
XX CC animal or in the manufacture of a medicament for the prophylactic
XX CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
XX CC or Shigella infection. The peptide that is immobilized on a solid support
XX CC is also useful as a reagent for determining the ability of a plasma
XX CC protein to bind to bacteria. The present sequence represents an E. coli
XX CC 15 kDa protein
XX
XX Sequence 151 AA;
XX
XX Query Match 66.4%; Score 518; DB 7; Length 151;
XX Best Local Similarity 68.2%; Pred. No. 3.8e-42;
XX Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNNGSSGPDYDQLVTRVVTHEMAHALQ 60
XX Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNNGSSGPDYDQLVTRVVTHEMAHALQ 60
XX
XX QY 61 SDARKSETTITQSGYGNAGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
XX Db 61 TDARNSDLTITQGGNGADVCGGSDSSIDLITQRFNGSATLDQNGKNSMTVKQFGG 120
XX
XX QY 121 NNAALVNQTASDSSVVMVQVGFNNATANQY 151
XX Db 121 GNGAAVDQTASNSSVNVTVQVGFNNATAHQY 151
XX
XX Search completed: August 2, 2004, 14:48:25
XX Job time : 44.9 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 686 | 87.9 | 151 | 1 | US-08-233-788A-59 |
| 2 | 506 | 64.9 | 120 | 1 | US-08-233-788A-57 |
| 3 | 92 | 11.8 | 975 | 4 | US-09-328-352-4764 |
| 4 | 90 | 11.5 | 906 | 1 | US-08-254-573-2 |
| 5 | 90 | 11.5 | 906 | 1 | US-08-687-379-2 |
| 6 | 90 | 11.5 | 906 | 1 | US-08-687-379-4 |
| 7 | 90 | 11.5 | 906 | 4 | US-08-172-332-1 |
| 8 | 90 | 11.5 | 906 | 4 | US-08-216-348-2 |
| 9 | 87.5 | 11.2 | 738 | 3 | US-08-864-038A-3 |
| 10 | 84 | 10.8 | 892 | 4 | US-09-336-447A-5 |
| 11 | 84 | 10.8 | 907 | 1 | US-07-718-575-2 |
| 12 | 84 | 10.8 | 907 | 1 | US-08-481-206-2 |
| 13 | 84 | 10.8 | 907 | 2 | US-08-486-269A-2 |
| 14 | 84 | 10.8 | 943 | 4 | US-09-056-556-204 |
| 15 | 84 | 10.8 | 943 | 4 | US-09-072-596-199 |
| 16 | 84 | 10.8 | 943 | 4 | US-09-477-135A-131 |
| 17 | 84 | 10.8 | 943 | 4 | US-09-072-967-204 |
| 18 | 84 | 10.8 | 1415 | 4 | US-09-253-991A-26438 |
| 19 | 83.5 | 10.7 | 415 | 4 | US-09-025-769B-280 |
| 20 | 83.5 | 10.7 | 873 | 4 | US-09-336-447A-13 |
| 21 | 83 | 10.6 | 518 | 3 | US-09-043-123-2 |
| 22 | 83 | 10.6 | 1912 | 1 | US-08-409-995-4 |
| 23 | 83 | 10.6 | 1912 | 3 | US-08-685-467-4 |
| 24 | 83 | 10.6 | 2353 | 3 | US-09-377-155-33 |
| 25 | 83 | 10.6 | 2353 | 4 | US-08-913-942-4 |
| 26 | 83 | 10.6 | 2353 | 4 | US-09-669-974-33 |
| 27 | 83 | 10.6 | 2353 | 4 | US-09-797-862-33 |

ALIGNMENTS

RESULT 1

US-08-233-788A-59
; Sequence 59, Application US/08233788A

; Patent No. 5635617

; GENERAL INFORMATION:

; APPLICANT: Doran, James L.

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, Karen S.

; APPLICANT: Clouthier, Sharon C.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

; TITLE OF INVENTION: OF SALMONELLA

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,788A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 59:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 151 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-233-788A-59

Query Match 87.9%; Score 686; DB 1; Length 151;

Best local Similarity 90.1%; Pred. No. 2.2e-62;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGGNSGDPYDQLVTRVVTTHMAHALQ 60

||||| : : : |||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALAQ 60
 QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
 Db 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
 Db 121 NNPALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 2
 US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouthier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 822-4900
 ; TELEFAX: (206) 822-6031
 ; TELEX: 3723836 SEEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.9%; Score 506; DB 1; Length 120;
 Best Local Similarity 87.5%; Pred. No. 3, 4e-44;
 Matches 98; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 22 VVPQWGGGNGHNGSSGPDYDQVLTVRVTHMAHALQSDARKSETTITQSGYNGADV 81
 Db 1 VVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADV 60
 QY 82 GCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGNNALVNQASPS 133
 Db 61 GCGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGNNALVNQASPS 112

RESULT 3
 US-09-328-352-4764
 ; Sequence 4764, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 4764
 ; LENGTH: 975
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-4764

Query Match 11.8%; Score 92; DB 4; Length 975;
 Best Local Similarity 23.8%; Pred. No. 0.75;
 Matches 35; Conservative 24; Mismatches 52; Indels 36; Gaps 7;
 QY 15 SGSALAGVVPQWGGGNGHNGG-QNSSGPDYDQVLTVRVTHM-----AHALQSDA 63
 Db 300 AGNGIA-----SCNGEHNYGIGNGDDVD--ITAPITGVNLISGNSFTLIGNSSSSV 351
 QY 64 RKSETTITQS-----GYNGADVGCGADNSTIETQNGF-----RNNATIDQWNAKNS 111
 Db 352 NTAPTTISNTVNDNTIDNGSGTSGSGNGSGDGLLNGAASNGEHNYGIGNGDDVD 411
 QY 112 DIT-----VGQYGGNNAALVNQASPS 133
 Db 412 DITSPITGIFNFGSGFSLIGNSSSS 438

RESULT 4
 US-08-254-573-2
 ; Sequence 2, Application US/08254573
 ; Patent No. 5610032
 ; GENERAL INFORMATION:
 ; APPLICANT: KAMBOJ, Rajender
 ; APPLICANT: ELLIOTT, Candace
 ; APPLICANT: NUTT, Stephen
 ; TITLE OF INVENTION: AMPA-BINDING HUMAN GluR1 RECEPTORS
 ; NUMBER OF SEQUENCES: 5
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 1800 Diagonal Road, Suite 500
 ; CITY: Alexandria
 ; STATE: Virginia
 ; COUNTRY: USA
 ; ZIP: 22313-0299
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/254,573
 ; FILING DATE: 06-JUN-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/896,611
 ; FILING DATE: 10-JUN-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 16777/179 ALLE
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 836-9300
 ; TELEFAX: (703) 883-4109
 ; TELEX: 899149
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 906 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein


```

; APPLICANT: Snyder, Yvonne M
; APPLICANT: Sharp, Robert L
; TITLE OF INVENTION: HUMAN GLUTAMATE RECEPTOR AND RELATED DNA
; TITLE OF INVENTION: COMPOUNDS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lilly Patent Division/JPL
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/172,332
; FILING DATE: 22-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/879,688
; FILING DATE: May 1, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Leeds, James P.
; REGISTRATION NUMBER: 35241
; REFERENCE/DOCKET NUMBER: X-8342
; TELEPHONE: 317-276-1667
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-172-332-1

```

```

Query Match 11.5%; Score 90; DB 4; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVSGSALAGVWPQWGGGNGGNSGGPDYDQL-----VTRVVTHE---M 55
DB 250 VTGFQLVNYDTIPAKIMQW-----KNSDARDHTRVDWKRPKYSALTIDGVKVM 300

QY 56 AHALQSDARKSETTTTQSGYNGADV-----GQGD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVPWQGGIDIQALQVRFEGLTGNVQFNE 357
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVPWQGGIDIQALQVRFEGLTGNVQFNE 357
QY 95 NGRNNAT-----IDQNAKNSDI---TVGQYGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTHLVEMKHDGIRKGYWNEDDKFVPAATDAQAGDSSVQNRITYVTITLED 417
QY 133 SSVVMRVQVFGNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

```

```

RESULT 8
US-08-216-326-2
; Sequence 2, Application US/08216326
; Patent No. 6406868
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: ELIOTT, Candace
; APPLICANT: NUTT, Stephen
; TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street N.W., Suite 500
; CITY: Washington, D.C.

```

```

; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,326
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,611
; FILING DATE: 10-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/229/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-216-326-2

```

```

Query Match 11.5%; Score 90; DB 4; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVSGSALAGVWPQWGGGNGGNSGGPDYDQL-----VTRVVTHE---M 55
DB 250 VTGFQLVNYDTIPAKIMQW-----KNSDARDHTRVDWKRPKYSALTIDGVKVM 300

QY 56 AHALQSDARKSETTTTQSGYNGADV-----GQGD-----NSTIELTQ 94
DB 301 AEAFOQLRRQ---RIDISRRGNAGDCLANPAVPWQGGIDIQALQVRFEGLTGNVQFNE 357
QY 95 NGRNNAT-----IDQNAKNSDI---TVGQYGNNAALVNQT-----ASD 132
DB 358 KGRRTNYTHLVEMKHDGIRKGYWNEDDKFVPAATDAQAGDSSVQNRITYVTITLED 417
QY 133 SSVVMRVQVFGNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

```

```

RESULT 9
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Ishinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95

```

```

; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
;
; US-08-864-038A-3
;
; Query Match 11.2%; Score 87.5; DB 3; Length 738;
; Best Local Similarity 25.6%; Pred. No. 1.5;
; Matches 41; Conservative 11; Mismatches 59; Indels 49; Gaps 5;
;
; QY 3 LLKVAFAAIVVSSALAGVVPQGGGNGHNGGNSGPDYDQLVTRVTHEMAHALQSD 62
; DB 419 LLKSASASASASASAG-----GGGGGNGGNGGGG-----GG 455
;
; QY 63 ARKSETTTTQSGYNGADVGGADNSTITLTQ-----NGFRNNATIDQWNAKNSDI 117
; DB 456 AGALAAALAAAGAGGLGGGGGGAALAAAGAGGGGGLGGL-----GG 503
;
; QY 118 YGNNALVNQTSDDS-----VMRVQVFGNNATA 148
; DB 504 LGGGSAALAAAAAAGGGGGRALRRLRQMRGGGSA 543
;
; RESULT 10
; US-03-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY-024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
;
; US-09-336-447A-5
;
; Query Match 10.8%; Score 84; DB 4; Length 892;
; Best Local Similarity 26.8%; Pred. No. 4.3;
; Matches 41; Conservative 17; Mismatches 55; Indels 40; Gaps 9;
;
; QY 28 GGNHNGGNS--GPDYDQLVTRVTHEMAHALQSDARKSETTI-----TQSGYNGAD 80
; DB 75 GGDNEAKGNYSTVGGDYNEAKGNYST--VGGGSSNTAKGKSTGGGDTNDANGTYST 132
;
; QY 81 VGG-----ADNSTI-----ELTQGFNNATIDQWNAKNSDI 120
; DB 133 IGGYYSRAIGDSSITIGGYNQATCEKSTVAGGRN-----QATGNNSTVAGGSYNQATG 188
;
; QY 121 NNAALV-----NOTASDSSVMVQVFGNNATAN 149
; DB 189 NNSTVAGGSHNQATGEGSF---AAGVENKANAN 218
;
; RESULT 11
; US-07-718-575-2
; Sequence 2, Application US/07718575
; Patent No. 5202257
; GENERAL INFORMATION:
; APPLICANT: Heinemann Ph.D., Stephen F.
; APPLICANT: Boulter Ph.D., James R.
; APPLICANT: Hollmann Ph.D., Michael NMN
; APPLICANT: Settler Ph.D., Bernhard NMN
; APPLICANT: Jensen Ph.D., Jan E.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
; METHOD OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 So. Flower St., Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: United States
; ZIP: 90071-2921
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/718,575
; FILING DATE: 19910813
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter Ph.D., Stephen E.
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: P31 8962
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; TELEX: 9103330318
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-07-718-575-2
;
; Query Match 10.8%; Score 84; DB 1; Length 907;
; Best Local Similarity 22.6%; Pred. No. 4.4;
; Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;
;
; QY 6 VAAFAAIVVSGSALAGVVPQGGGNGHNGG--NSGPDYDQLVTRVTHEMAHALQSDAR 64
; DB 250 VTGFQLVNVTDTIPARIMQWRTSDSRDHTRVDWKRPKYTSALTVDGVKVAEAFQSLRR 309
;
; QY 65 KSETTITQSGYNGADV-----GQAD-----NSTIETQGFNNAT- 102
; DB 310 Q---RIDISRRGNAGDCLANPAVPGQGGIDIORALQOVFEGLTGNVQNEKGRNRYTL 366
;
; QY 103 -----IDQWNAKNSDI---TVQYGGNNAALVNQT-----ASDSSVMVQV 141
; DB 367 HVTEMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRIVTVITILEDPYVWLKX-- 424

```

QY 142 FGNATANOY 151
Db 425 -----NANQF 429

RESULT 12
US-08-481-206-2
; Sequence 2, Application US/08481206
; Patent No. 5739291
; GENERAL INFORMATION:
; APPLICANT: Heinemann Ph.D., Stephen P.
; APPLICANT: Boulter Ph.D., James R.
; APPLICANT: Hollmann Ph.D., Michael NNN
; APPLICANT: Bettler Ph.D., Bernhard NNN
; APPLICANT: Jensen Ph.D., Jan E.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 So. Flower St., Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: United States
; ZIP: 90071-2921
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,206
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/013,767
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter Ph.D., Stephen E.
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: P31 8962
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; TELEX: 9103330318
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-481-206-2

Query Match 10.8%; Score 84; DB 1; Length 907;
Best Local Similarity 22.6%; Pred. No. 4.4;
Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;
QY 6 VAAFAIVVSGSALAGVVPQWGGGNHNGG-NSSGPDYDQLVTRVVTHEMAHALQSDAR 64
Db 250 VTGFQLVNYTDTIPARIMQWRTSDRHTRDVWKRPKYTSALTYDGVKVMAEAFQSLRR 309
QY 65 KSETTITQSGYNGADV-----GQAD-----NSTIELTQNGPRNAT- 102
Db 310 Q---RIDISRRNAGDCLANPAVPWGQGDIDQALQQVRFEGLTGNVQFNEKGRNTYTL 366
QY 103 -----IDWNKNSDI-----TVQYGGNNAALVNQT-----ASDSSVMVRQVG 141
Db 367 HVIEMKHDGIRKIGYWNEDDKFVPAATDAQAGDGNSSVQNRITYIVTTILEDPPYMLKK-- 424
QY 142 FGNATANOY 151
Db 425 -----NANQF 429

RESULT 13
US-08-486-269A-2
; Sequence 2, Application US/08486269A
; Patent No. 5945509
; GENERAL INFORMATION:
; APPLICANT: Heinemann, Stephen P.
; APPLICANT: Boulter, James R.
; APPLICANT: Hollmann, Michael
; APPLICANT: Bettler, Bernhard
; APPLICANT: Jensen, Jan E.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,269A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/013,767
; FILING DATE: 04-FEB-1993
; APPLICATION NUMBER: 07/718,575
; FILING DATE: 21-JUN-1991
; APPLICATION NUMBER: PCT/US90/06153
; FILING DATE: 25-OCT-1990
; APPLICATION NUMBER: 07/428,116
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9986
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-486-269A-2

Query Match 10.8%; Score 84; DB 2; Length 907;
Best Local Similarity 22.6%; Pred. No. 4.4;
Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;
QY 6 VAAFAIVVSGSALAGVVPQWGGGNHNGG-NSSGPDYDQLVTRVVTHEMAHALQSDAR 64
Db 250 VTGFQLVNYTDTIPARIMQWRTSDRHTRDVWKRPKYTSALTYDGVKVMAEAFQSLRR 309
QY 65 KSETTITQSGYNGADV-----GQAD-----NSTIELTQNGPRNAT- 102
Db 310 Q---RIDISRRNAGDCLANPAVPWGQGDIDQALQQVRFEGLTGNVQFNEKGRNTYTL 366
QY 103 -----IDWNKNSDI-----TVQYGGNNAALVNQT-----ASDSSVMVRQVG 141
Db 367 HVIEMKHDGIRKIGYWNEDDKFVPAATDAQAGDGNSSVQNRITYIVTTILEDPPYMLKK-- 424

QY 142 FGNATANOY 151
Db 425 -----NANQF 429

RESULT 14

US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-204

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 25.2%; Pred. No. 4.7;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;
QY 15 SGSALAGVVPQWGGGNGHNGSGDPDYDLVTRVVTHEMAHALQSDARKSETTITQSG 74
Db 184 SGTGNVGI-----GNSGTGNWIGNSGNSYN-----TG 211
QY 75 YNGADVCGGADNSTIELTQNGFRNATIDQNAKNSD---ITVGOY-----GGNAA 124
Db 212 FGNSGDANTGFNSGIANTGVGNAGNYNTGSPNSNTGTFNMGQYNTGYLNSGNYNTG 271
QY 125 LVN 127
Db 272 LAN 274

RESULT 15

US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-199

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 25.2%; Pred. No. 4.7;
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;
QY 15 SGSALAGVVPQWGGGNGHNGSGDPDYDLVTRVVTHEMAHALQSDARKSETTITQSG 74
Db 184 SGTGNVGI-----GNSGTGNWIGNSGNSYN-----TG 211
QY 75 YNGADVCGGADNSTIELTQNGFRNATIDQNAKNSD---ITVGOY-----GGNAA 124
Db 212 FGNSGDANTGFNSGIANTGVGNAGNYNTGSPNSNTGTFNMGQYNTGYLNSGNYNTG 271
QY 125 LVN 127
Db 272 LAN 274

Search completed: August 2, 2004, 14:58:33
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATNQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------------|-------------------|
| 1 | 520 | 66.7 | 151 | 12 US-09-741-873B-4 | Sequence 4, Appli |
| 2 | 520 | 66.7 | 151 | 12 US-09-741-873B-4 | Sequence 4, Appli |
| 3 | 444 | 56.9 | 131 | 12 US-09-741-873B-2 | Sequence 2, Appli |
| 4 | 444 | 56.9 | 131 | 12 US-09-741-873B-2 | Sequence 2, Appli |
| 5 | 108.5 | 13.9 | 445 | 15 US-10-369-493-20638 | Sequence 20638, A |
| 6 | 93 | 11.9 | 974 | 12 US-10-282-122A-44999 | Sequence 44999, A |
| 7 | 91.5 | 11.7 | 580 | 12 US-10-647-057-4 | Sequence 4, Appli |
| 8 | 91 | 11.7 | 597 | 9 US-09-793-306-146 | Sequence 146, App |
| 9 | 90 | 11.5 | 906 | 12 US-10-233-449-7 | Sequence 7, Appli |
| 10 | 90 | 11.5 | 906 | 14 US-10-251-661-2 | Sequence 2, Appli |
| 11 | 89 | 11.4 | 1448 | 16 US-10-408-765A-998 | Sequence 998, App |
| 12 | 88.5 | 11.3 | 204 | 12 US-10-424-599-203972 | Sequence 203972, |
| 13 | 88.5 | 11.3 | 244 | 12 US-10-425-114-37715 | Sequence 37715, A |
| 14 | 88.5 | 11.3 | 253 | 16 US-10-437-963-114193 | Sequence 114193, |
| 15 | 88.5 | 11.3 | 283 | 12 US-10-424-599-232271 | Sequence 232271, |

| | | | | | |
|----|------|------|------|-------------------------|--------------------|
| 16 | 88 | 11.3 | 3619 | 12 US-10-282-122A-67768 | Sequence 67768, A |
| 17 | 88 | 11.3 | 6310 | 12 US-10-282-122A-67793 | Sequence 67793, A |
| 18 | 87.5 | 11.2 | 172 | 12 US-10-425-114-46013 | Sequence 46013, A |
| 19 | 86.5 | 11.1 | 208 | 16 US-10-475-970-16 | Sequence 16, Appli |
| 20 | 85.5 | 11.0 | 278 | 9 US-09-810-264-28 | Sequence 28, Appli |
| 21 | 85.5 | 11.0 | 588 | 12 US-10-282-122A-64869 | Sequence 64869, A |
| 22 | 84.5 | 10.8 | 408 | 15 US-10-369-493-12833 | Sequence 12833, A |
| 23 | 84.5 | 10.8 | 847 | 16 US-10-437-963-118741 | Sequence 118741, |
| 24 | 84 | 10.8 | 283 | 12 US-10-424-599-168890 | Sequence 168890, |
| 25 | 84 | 10.8 | 892 | 10 US-09-952-267-5 | Sequence 5, Appli |
| 26 | 84 | 10.8 | 943 | 9 US-09-996-634-131 | Sequence 131, App |
| 27 | 84 | 10.8 | 943 | 10 US-09-997-182-131 | Sequence 131, App |
| 28 | 84 | 10.8 | 943 | 10 US-09-997-181-131 | Sequence 131, App |
| 29 | 84 | 10.8 | 943 | 14 US-10-193-002-199 | Sequence 199, App |
| 30 | 84 | 10.8 | 943 | 14 US-10-084-843-204 | Sequence 204, App |
| 31 | 84 | 10.8 | 1246 | 12 US-10-282-122A-49773 | Sequence 49773, A |
| 32 | 84 | 10.8 | 3300 | 12 US-10-282-122A-64369 | Sequence 64369, A |
| 33 | 83.5 | 10.7 | 65 | 9 US-09-996-194-16 | Sequence 16, Appli |
| 34 | 83.5 | 10.7 | 65 | 12 US-10-164-966-33 | Sequence 33, Appli |
| 35 | 83.5 | 10.7 | 628 | 12 US-10-282-122A-53269 | Sequence 53269, A |
| 36 | 83.5 | 10.7 | 873 | 10 US-09-953-267-13 | Sequence 13, Appli |
| 37 | 83 | 10.6 | 182 | 14 US-10-238-075-1549 | Sequence 1549, Ap |
| 38 | 83 | 10.6 | 354 | 10 US-09-820-843A-21 | Sequence 21, Appli |
| 39 | 83 | 10.6 | 518 | 9 US-09-976-297-2 | Sequence 2, Appli |
| 40 | 83 | 10.6 | 720 | 9 US-09-801-368-176 | Sequence 176, App |
| 41 | 83 | 10.6 | 961 | 12 US-10-282-122A-76322 | Sequence 76322, A |
| 42 | 83 | 10.6 | 1306 | 12 US-10-282-122A-64405 | Sequence 64405, A |
| 43 | 83 | 10.6 | 2353 | 9 US-09-797-862-33 | Sequence 33, Appli |
| 44 | 82.5 | 10.6 | 434 | 15 US-10-045-674-594 | Sequence 594, App |
| 45 | 82.5 | 10.6 | 574 | 12 US-10-282-122A-55429 | Sequence 55429, A |

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4; Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Olsear, Arne
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1994-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.7%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVVPQWGGGNGHNGGSSGPDYDQIVTVRVVTHEMAHLQ 60


```
Best Local Similarity 65.6%; Pred. No. 1.2e-37;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

QY 21 GWFPQWGGGNNHGGSSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80
Db 1 GWFPQWGGGNNHGGSSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 60

QY 81 VGGADNSTIELTQGFNNATIDQWNAKSDITVGGYGGNNAALVNQTSDDSSVMYROY 140
Db 61 VGGSDSSSIDLTQGFNSATLDQWNGKNSMTVKQFGGNGAAYDQTSNSSVNTQV 120

QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.9%; Score 108.5; DB 15; Length 445;
Best Local Similarity 27.5%; Pred. No. 0.015;
Matches 46; Conservative 19; Mismatches 61; Indels 41; Gaps 6;

QY 7 AAFRA-----IVVSGSALAGVVPQWGGG-----NHNG-----GNSSGPDYDQLVTRV 50
Db 19 AAFADSNTVYLNQDGNQANITQSGNGNSVGAFNGSGFLQNGTLSGAN---LLT-- 73

QY 51 VTHEMAHALQSDARKSETTITQSGYNGADYQGGADNSTIELTQGFNNATIDQWNAKN 110
Db 74 ---VKQSGNSNSVGRDITQKQSGAGNSAIFQEGTGSDELQQTGTSNGAVPSGWNWTN 129

QY 111 -----SDITVGGYGGNNAALVNQTSDDSSVMYRQVG 141
Db 130 DPGVFNKITQDSSNSGKSVIQDGKNVFSIKQGTGNSTSVNQIG 176

RESULT 6
US-10-282-122A-44999
; Sequence 44999, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 44999
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-10-282-122A-44999

Query Match 11.9%; Score 93; DB 12; Length 974;
Best Local Similarity 24.5%; Pred. No. 1.5;
Matches 36; Conservative 23; Mismatches 52; Indels 36; Gaps 7;

QY 15 SGSALAGVVPQWGGGNNHGG-GNSSGPDYDQLVTRVVTHEM-----AHALQSDA 63
Db 299 ANGIA-----SGNGEYHNGGNGDDVD--ITAPITGVNLISGNSFTLIGNSSSV 350

QY 64 RKSETTITQS-----GYGNGADVGGADNSTIELTQNGF-----RNNATIDQWNAKNS 111
Db 351 NTAPITTSNTVNDNTIDNGSGGTGSGNGSGDGLLNGAASGNGEYHNGGNGDDV 410

QY 112 DIT-----VGQYGGNNAALVNQTSADS 133
Db 411 DITAPITGVNFGNSFLIGNSSSV 437

RESULT 7
US-10-647-057-4
; Sequence 4; Application US/10647057
; Publication No. US20040047871A1
; GENERAL INFORMATION:
; APPLICANT: KANSAS STATE UNIVERSITY RESEARCH FOUNDATION
; APPLICANT: NAGARAJA, T.
; APPLICANT: STEWART, GEORGE
; APPLICANT: NARAYANAN, SANJEEV
; APPLICANT: CHENGAPPA, M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN VACCINE AND PRE
; FILE REFERENCE: PCT-30962
; CURRENT APPLICATION NUMBER: US/10/647,057
; CURRENT FILING DATE: 2003-08-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
```


QY 56 AHALQSDARKSETTITQSGYNGADV-----GQAD-----NSTIELTQ 94
Db 301 AEAFOSLRQ---RIDISRRGNAGDCLANPAVPWGQIDIGRALQQVAFELTGNVQFNE 357
QY 95 NGFRNAT-----IDWNKNSDI---TVQYGGNNAALVNQT-----ASP 132
Db 358 KGRNTYTLHVIEKHGIRKIGYWNEDKFPATDAQAGDNSSVQNRITYIVTTILED 417
QY 133 SSVVMVRQVGFNNATANY 151
Db 418 PYWMLK-----NANQF 429

RESULT 11

US-10-408-765A-998
; Sequence 998, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 998
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-998

Query Match 11.4%; Score 89; DB 16; Length 1448;
Best Local Similarity 24.4%; Pred. No. 6.5;
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;
QY 10 AAIVVSGSALAGV-----PWG-----GGN-----H 32
Db 307 AAAAKSGHAWGAAANQEDKSPWGPBPKPSQHWGQGRSNPAKSGGDWADSSVLGH 366
QY 33 NCGGNSGPDYDQLVTRVVTHEMAHALQSDARKSETTIT-OSGYNGADVCGGADNSTIE 91
Db 367 LGDGKKGSGMD-----ADSNRSGGWNDDTTRSGNSGWNSTNTKANPCTNWGE 415
QY 92 LTQNGFRNATIDWNKNSDITVQYGGNNAALVNQTAS 131
Db 416 TLKPGPQN-----NASKFPQDNVSNWGG--AASVKQTGT 448

RESULT 12

US-10-424-599-203972
; Sequence 203972, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 203972
; LENGTH: 204
; TYPE: PRT

; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_26213C.1.pap
US-10-424-599-203972
Query Match 11.3%; Score 88.5; DB 12; Length 204;
Best Local Similarity 25.7%; Pred. No. 0.61; Indels 61; Gaps 7;
Matches 35; Conservative 10; Mismatches 30; Indels 61; Gaps 7;
QY 21 GVVFPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80
Db 106 GRVP-W-GGNSHERGGYSDG-----NSD-----SGWQGGQD 134
QY 81 VGOQADNSTIELTQNGFRNATIDWNKNSDIT-----VGOYGGNNAALVNQTAS 131
Db 135 QGSGSDG-----NGQGRGWRGNGNNSERNKDEESRGNSGWSGFGNAGSGNENS- 182
QY 132 DSSVMVRQVGFNNAT 147
Db 183 -----GWGKNAT 189
RESULT 13
US-10-425-114-37715
; Sequence 37715, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 37715
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3170-053-G12_FLI.pap
US-10-425-114-37715
Query Match 11.3%; Score 88.5; DB 12; Length 244;
Best Local Similarity 22.9%; Pred. No. 0.76; Indels 39; Gaps 6;
Matches 32; Conservative 21; Mismatches 48; Indels 39; Gaps 6;
QY 27 GGGGNGHNGGNSGPDY-----DQLVTRVVTHEMAHALQSDARKSETTITQSG- 74
Db 23 GGGGNNNNNNNEGTYNFATSSAPSLFSNSVNASALSHMSATALLQKAAQMGATTNGGT 82
QY 75 -----YNGADVGOQADNSTIELTQNGFRNATIDWNKNSDITVGOYGGNNAALVN- 127
Db 83 ASLLKFSGSASSSSGGGSKLV-----NAAAN--YVSGMFGNNH--VNE 122
QY 128 QTASDSSVMVRQVGFNNAT 147
Db 123 QSNLNQDLNNSFAVGNS 142

RESULT 14

US-10-437-963-114193
; Sequence 114193, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

Search completed: August 2, 2004, 15:36:11
Job time : 37.8 secs

APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 114193
LENGTH: 253
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_17908C.1.pap
US-10-437-963-114193

Query Match 11.3%; Score 88.5; DB 16; Length 253;
Best Local Similarity 26.6%; Pred. No. 0.8;
Matches 38; Conservative 14; Mismatches 48; Indels 43; Gaps 6;

QY 4 LKVA--RAIVVSSGALAGVVPWG-----GGGNHGGG-----NSSGP 41
DB 65 IKLALGFVLLSIGLASARVERYSSESGGTNGEGGVYNGGVSGSGAGSGSGG 124
QY 42 DYDQLVTRVVTHEMAHALQSDARKSETTITQSGY----GNGADVCGGADNSTIELTQNGF 97
DB 125 NYG-----AHASGGGGEG-----GGYQYGGSGSGSGSGSGSSEYQTNGG 167
QY 98 RNNATIDQNAKNSDITVQYGG 120
DB 168 YGYAGSSSAGSGGAGAGAGGAGG 190

RESULT 15
US-10-424-599-232271
Sequence 232271, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 232271
LENGTH: 283
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_51764C.1.pap
US-10-424-599-232271

Query Match 11.3%; Score 88.5; DB 12; Length 283;
Best Local Similarity 22.9%; Pred. No. 0.92;
Matches 32; Conservative 21; Mismatches 48; Indels 39; Gaps 6;

QY 27 GGGNHGGGSGSPDY-----DQLVTRVVTHEMAHALQSDARKSETTITQSG- 74
DB 62 GGGNNNNNNNEGTYFATSSAPSLFSNVNASALSHMSATALLQKAAQMGATTSNGGT 121
QY 75 -----YNGADVGGGADNSTIELTQNGFRNATIDQNAKNSDITVQYCGNNAALVN- 127
DB 122 ASLLKSFSGASSSGSGGSKLV-----NAAN--YVGMFGGNH--VNE 161
QY 128 QTASDSVMVQVGFQNNAT 147
DB 162 QSNLNQLDLNLSFVGGKSS 181

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFGNATANYQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 97603577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Parents AA Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US105_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
|------------|-------|-------------|--------|-------|-------------|

| | | | | | | |
|----|-------|-------|------|----|---------------------|-------------------|
| 1 | 780 | 100.0 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 2 | 691 | 87.6 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 3 | 686 | 87.9 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 4 | 675 | 86.5 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 5 | 622 | 79.7 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 6 | 618 | 79.2 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 7 | 613 | 78.6 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 8 | 611 | 78.3 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 9 | 608 | 77.9 | 151 | 19 | US-09-543-407-34 | Sequence 24, Appl |
| 10 | 604 | 77.4 | 131 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 11 | 601 | 77.1 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 12 | 600 | 76.9 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 13 | 599 | 76.8 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 14 | 523 | 67.1 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 520 | 66.7 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 520 | 66.7 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 518 | 66.4 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 518 | 66.4 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 506 | 64.9 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 483 | 61.9 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 21 | 457 | 58.6 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 457 | 58.6 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 23 | 457 | 58.6 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 24 | 444 | 56.9 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 444 | 56.9 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 359.5 | 46.1 | 109 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 27 | 276 | 35.4 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 28 | 250.5 | 32.1 | 70 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 29 | 175 | 22.4 | 48 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 30 | 116 | 14.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 31 | 116 | 14.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 116 | 14.9 | 186 | 30 | US-10-417-886-5833 | Sequence 5833, Ap |
| 33 | 108.5 | 13.9 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 34 | 108.5 | 13.9 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 35 | 107.5 | 13.8 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 36 | 106.5 | 13.7 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 37 | 101 | 12.9 | 257 | 33 | US-60-173-464-21553 | Sequence 21553, A |
| 38 | 101 | 12.9 | 262 | 20 | US-09-614-150-25818 | Sequence 25818, A |
| 39 | 101 | 12.9 | 262 | 20 | US-09-614-150-25818 | Sequence 25818, A |
| 40 | 101 | 12.9 | 262 | 33 | US-60-191-637-25957 | Sequence 25957, A |
| 41 | 101 | 12.9 | 262 | 33 | US-60-191-637-25957 | Sequence 25957, A |
| 42 | 100 | 12.8 | 1308 | 27 | US-10-179-131-5148 | Sequence 5148, Ap |
| 43 | 99.5 | 12.8 | 1249 | 30 | US-10-455-719-358 | Sequence 358, App |
| 44 | 99.5 | 12.8 | 1249 | 33 | US-60-385-568-357 | Sequence 357, App |
| 45 | 99.5 | 12.8 | 1249 | 33 | US-60-446-775-358 | Sequence 358, App |

ALIGNMENTS

RESULT 1
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.


```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      78.3%; Score 611; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.5e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDYDQLVTRVVTHEMAHALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120

QY 121 NNAALVNQTASDSSVMVROVQFGNNATANQY 151
DB 121 LVTRVVVTHEMAHASVMVROVQFGNNATANQY 151

RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      77.9%; Score 608; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.2e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDYDQLVTRVVTHEMAHALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVVT 120
```

```

QY 121 NNAALVNQTASDSSVMVROVQFGNNATANQY 151
DB 121 HEMAHANQTASDSSVMVROVQFGNNATANQY 151

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
; OTHER INFORMATION: Recombinant Salmonella enteritidis
US-09-543-407-31

Query Match      77.4%; Score 604; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 2.6e-56;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPQWGGGNGHNGSGGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60

QY 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGNNAAALVNQTASDSSVMVROV 140
DB 61 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGNNAAALVNQTASDSSVMVROV 120

QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match      77.1%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.6e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGGPDYDQLVTRVVTHEMAHALQ 60
```

```
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
```

RESULT 12

```
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22
```

```
Query Match 76.9%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 8.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
```

RESULT 13

```
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 76.8%; Score 599; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.1e-55;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
```

RESULT 14

```
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7
```

```
Query Match 67.1%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-47;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 TDARNSDLTITQHGNGGADYVQGGSDSDSIDTQGRGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NGGAADVDTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 15

```
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match          66.7%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.4e-47;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDYDOLVTRVVTHEMAHALQ 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 MKLLKVAATAAIVFSGSAVAGVVPQYGGGNNHGGSSGPNSELNITYYGGGNSALALQ 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 61 SDARKSETTITOSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNNAKNSDITVGOYGG 120
   :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db 61 TDARNSDLTITQGGGNGADVCGGDDSSIDLTPQFGNSATLDQNGKNSMTVKQFEGG 120
   :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY 121 NNAALVNOTASDSVMVRQVGFNNATANQY 151
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: August 2, 2004, 15:26:43
Job time : 167.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pcp.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pcp.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pcp.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pcp.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pcp.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pcp.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 520 | 66.7 | 151 | 5 | US-09-741-873C-4 |
| 2 | 444 | 56.9 | 131 | 5 | US-09-741-873C-2 |
| 3 | 90 | 11.5 | 852 | 6 | US-10-170-205E-3413 |
| 4 | 89 | 11.4 | 1627 | 6 | US-10-170-205E-16659 |
| 5 | 89 | 11.4 | 1905 | 1 | PCT-US04-09388-9 |
| 6 | 87 | 11.2 | 461 | 7 | US-60-556-841-11622 |
| 7 | 86 | 11.0 | 841 | 7 | US-60-565-632-7906 |
| 8 | 86 | 11.0 | 841 | 7 | US-60-579-062-7906 |
| 9 | 85 | 10.9 | 295 | 6 | US-10-425-115-312468 |
| 10 | 84 | 10.8 | 892 | 5 | US-09-952-267B-5 |
| 11 | 84 | 10.8 | 892 | 6 | US-10-872-768-5 |
| 12 | 84 | 10.8 | 892 | 6 | US-10-872-769-5 |
| 13 | 84 | 10.8 | 1507 | 7 | US-60-579-902-6263 |
| 14 | 83.5 | 10.7 | 415 | 5 | US-09-490-324-280 |
| 15 | 83.5 | 10.7 | 873 | 5 | US-09-952-267B-13 |
| 16 | 83.5 | 10.7 | 873 | 6 | US-10-872-768-13 |
| 17 | 83.5 | 10.7 | 873 | 6 | US-10-872-769-13 |
| 18 | 83 | 10.6 | 358 | 5 | US-09-248-796A-22578 |
| 19 | 83 | 10.6 | 556 | 6 | US-10-425-115-337674 |
| 20 | 83 | 10.6 | 573 | 7 | US-60-565-632-7907 |
| 21 | 83 | 10.6 | 573 | 7 | US-60-579-062-7907 |
| 22 | 82.5 | 10.6 | 434 | 6 | US-10-045-674A-594 |
| 23 | 81.5 | 10.4 | 177 | 6 | US-10-737-290-171 |
| 24 | 81.5 | 10.4 | 400 | 6 | US-10-490-953-13 |
| 25 | 81.5 | 10.4 | 400 | 6 | US-10-490-953-14 |
| 26 | 81.5 | 10.4 | 400 | 6 | US-10-490-953-20 |

| | | | | | | |
|----|------|------|------|---|----------------------|-------------------|
| 27 | 81.5 | 10.4 | 400 | 6 | US-10-723-981-14 | Sequence 14, Appl |
| 28 | 81.5 | 10.4 | 400 | 6 | US-10-723-981-15 | Sequence 15, Appl |
| 29 | 81.5 | 10.4 | 424 | 6 | US-10-045-674A-591 | Sequence 591, App |
| 30 | 81.5 | 10.4 | 533 | 6 | US-10-045-674A-527 | Sequence 527, App |
| 31 | 81.5 | 10.4 | 599 | 6 | US-10-491-001-22 | Sequence 22, Appl |
| 32 | 81 | 10.4 | 201 | 6 | US-10-425-115-309662 | Sequence 309662, |
| 33 | 81 | 10.4 | 412 | 7 | US-60-565-632-7905 | Sequence 7905, Ap |
| 34 | 81 | 10.4 | 412 | 7 | US-60-579-062-7905 | Sequence 7905, Ap |
| 35 | 80.5 | 10.3 | 258 | 6 | US-10-425-115-300390 | Sequence 300390, |
| 36 | 80.5 | 10.3 | 586 | 1 | PCT-US03-24982A-317 | Sequence 317, App |
| 37 | 80 | 10.3 | 132 | 6 | US-10-425-115-351875 | Sequence 351875, |
| 38 | 80 | 10.3 | 163 | 5 | US-09-490-324-282 | Sequence 282, App |
| 39 | 80 | 10.3 | 197 | 6 | US-10-425-115-304391 | Sequence 304391, |
| 40 | 80 | 10.3 | 374 | 1 | PCT-US04-11210-37 | Sequence 37, Appl |
| 41 | 80 | 10.3 | 2319 | 1 | PCT-US04-12717-26 | Sequence 26, Appl |
| 42 | 80 | 10.3 | 2319 | 6 | US-10-831-070-26 | Sequence 26, Appl |
| 43 | 79.5 | 10.2 | 376 | 6 | US-10-491-733-2 | Sequence 2, Appl |
| 44 | 79.5 | 10.2 | 956 | 6 | US-10-093-037A-63 | Sequence 63, Appl |
| 45 | 79 | 10.1 | 284 | 1 | PCT-US04-11210-36 | Sequence 36, Appl |

ALIGNMENTS

RESULT 1
US-09-741-873C-4
; Sequence 4: Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.7%; Score 520; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.6e-39;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| Qy | 1 | MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSSGPDYDQLVTRVWTHMAHALQ | 60 |
| Db | 1 | MKLLKVAFAAIVFSGSAVAGVVPQGGGNGGNGSGPNSLNIVYGGNSALALQ | 60 |
| Qy | 61 | SDARKSETTITGSGYNGADVGQADNSTIETQGFNRNATIDQWAKNSDITVGYGG | 120 |
| Db | 61 | TDARNSDLAITGCGNGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVQFGG | 120 |
| Qy | 121 | NNALYNQATSSSSVMVRQVGFNNATANQY | 151 |
| Db | 121 | NGAAVDQATSSSNVNTQVGFNNATAHOY | 151 |

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsep, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match          56.9%; Score 444; DB 5; Length 131;
Best Local Similarity 65.6%; Pred. No. 1.5e-32;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVPQWGGGNGHNGSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80
Db 1 GVPQYGGGNGHGGGNGSGPSELNTYQYGGGNSALALQTDARNSDLTITQHGCGNGAD 60
QY 81 VCGQADNSTIETONGFRNNATIDOWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROV 140
Db 61 VCGQDSDSIDITQSGFNGSATLDOWNKNSMTVQFGGNGAAGVDDGTASNSVNVTVQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 3
US-10-170-205E-3413
; Sequence 3413, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3413
; LENGTH: 852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-3413

Query Match          11.5%; Score 90; DB 6; Length 852;
Best Local Similarity 22.6%; Pred. No. 7.4;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVGSALAGVVPQWGGGNGHNGSGGPDYDQL-----VTRVVTHE-----M 55

```

```

Db 263 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGVKUM 313
QY 56 AHALQSDARKSETTITQSGYNGADV-----GQAD-----NSTIELTQ 94
Db 314 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVFWPGGIDIQRALQQRFEGLTGNVQFNE 370
QY 95 NGRFNAT-----IDWNKNSDI---TVGOYGGNNAALVNOT-----ASD 132
Db 371 KGRRTNYTLHVEMKHDGIRKGYWNEDDKFVPAATDAQAGGDNSSVQNRITVITILED 430
QY 133 SSVNRQVGFGNATANQY 151
Db 431 PYVWLKK-----NANQF 442

RESULT 4
US-10-170-205E-16659
; Sequence 16659, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16659
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-16659

Query Match          11.4%; Score 89; DB 6; Length 1627;
Best Local Similarity 24.4%; Pred. No. 19;
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;

QY 10 AAIWGSALAGVV-----PQW-----GGN-----H 32
Db 621 AAAAKSGHAWGAANDEKSPWTGPEPKPSQHWGDGQRSPAWAGGDWADSSVLGH 680
QY 33 NGGNSGGPDYDQLVTRVVTHEMAHALQSDARKSETTIT-OSGYNGADVGGADNSTIE 91
Db 681 LGGKNGSGWD-----ADNRSRGWNDITRSGNSGWNSTNTKANPGTNWGE 729
QY 92 LTQNGFRNNATIDOWNAKNSDITVGOYGGNNAALVNOTAS 131
Db 730 TLKPGPQON-----WASKPDNNVSNWGG--AASVKQTGT 762

RESULT 5
PCT-US04-09388-9
; Sequence 9, Application PC/TUS0409388
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; MARQUIS, Joseph P.;
; APPLICANT: TRAN, Uyen K.; YANG, Yonghong G.;
; APPLICANT: TANG, Y. Tom; CHAWLA, Narinder K.;
; APPLICANT: MURAGE, Jaji; WANG, Jonathan T.;
; APPLICANT: ELLIOTT, Vicki S.; CHIEN, David;
; APPLICANT: YUE, Henry; AZIMZAI, Yalda;
; APPLICANT: ISON, Craig H.; KHARE, Reena;
; APPLICANT: JIN, Pei; RAMKUMAR, Jayalaxmi;
; APPLICANT: FAVERO, Kristin D.; RICHARDSON, Thomas W.;
; APPLICANT: HAFALIA, April J.A.; BAUGHN, Mariah R.;
; APPLICANT: BECHA, Shanya D.; WILSON, Amy D.
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PF-1531 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/09388
; CURRENT FILING DATE: 2004-04-01
; PRIOR APPLICATION NUMBER: US 60/457,403
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/465,568
; PRIOR FILING DATE: 2003-04-24

```



```

; PRIOR APPLICATION NUMBER: US 60/476,135
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 60/476,583
; PRIOR FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 1905
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7525307CD1
PCT-US04-09388-9

Query Match      11.4%; Score 89; DB 1; Length 1905;
Best Local Similarity 24.4%; Pred. No. 23;
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;

QY      10 AAIVVSGSALAGV-----PWG-----GGGN-----H 32
Db      759 AAAAASKGAWGAANQEKSTWEPKPKSQHWGDCQSRNPASAGGDWADSSVLGH 818

QY      33 NCGGSSGPDYDQLVTRVVTHEMAHALQSPARKSETTIT-QSGYGNAGADVGGADNSTIE 91
Db      819 LGDGKKGKSGWD-----ADSNRSGGWNDITRSGNSGWNSTNTKANPGTNWGE 867

QY      92 LTQGNFRNATIDOWNAKNSDITVQYGGNNALVNTAS 131
Db      868 TLKPGPQON-----WASKPQDNVNSNMGG--AASVKQTGT 900

RESULT 6
US-60-556-841-11622
; Sequence 11622, Application US/60556841
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)
; CURRENT APPLICATION NUMBER: US/60/556,841
; CURRENT FILING DATE: 2004-03-25
; NUMBER OF SEQ ID NOS: 12463
; SEQ ID NO 11622
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Cytophaga hutchinsonii
US-60-556-841-11622

Query Match      11.2%; Score 87; DB 7; Length 461;
Best Local Similarity 24.2%; Pred. No. 6.7;
Matches 40; Conservative 22; Mismatches 69; Indels 34; Gaps 7;

QY      4 LKVAAP-AAIVVSGSALAGVPPVGGGGNHHGG---GNSSGPDYDQLVTRVVTHEMAHA 58
Db      177 VSVGAFHTAIKTDGSL-----WAGSNNGNGLGIGTITTKNAETKVGTATNWKSVYA 229

QY      59 LQSD--ARKSETTITQSGYGNAGDVGGADNSTIELTQGNFRNATIDOWN----- 107
Db      230 AGSNTFAIKTDGSLWAGVYNAVGLDGTITTDRLSPVQIGTDNN-----WKTISDGNST 284

QY      108 -AKNSDITVQYGGNNALVNTASDSSVMVQVGFGNNTANQY 151
Db      285 FALKSDGLUWGDNSNGQFGNGTITNTLPIQIG-----TDNQW 324

RESULT 7
US-60-565-632-7906
; Sequence 7906, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.

```


; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US/09/336,447
; FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-769-5

Query Match 10.8%; Score 84; DB 6; Length 892;
Best Local Similarity 26.8%; Pred. No. 27;
Matches 41; Conservative 17; Mismatches 55; Indels 40; Gaps 9;
QY 28 GGGHNGGCGSS---GPDYDQVTRVVTHEMAHALQSDARKSETTI---TQSGYNGAD 80
DB 75 GCKNEAKGNTVGGDYNEAKGNYSY---VGGSSNTAKEKSTIGGGTDANGYIST 132
QY 81 VQOG-----ADNSTI-----ELTQGFNNATIDQWNAKSDITVQYQY---G 120
DB 133 IGGGYISRAIGDSSTIGGGYNNQATGEKSTVAGGRNN---QATGNNSTVAGGSYNQATG 188
QY 121 NNAALV-----NOTASDSSVMVQVGFQGNNTAN 149
DB 189 NNSTVAGSHNQATGEGSF---AAGVENKANAN 218

RESULT 13
US-60-579-902-6263
; Sequence 6263, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; FILE REFERENCE: 38-21 (53444) A
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 6263
; LENGTH: 1507
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-6263

Query Match 10.8%; Score 84; DB 7; Length 1507;
Best Local Similarity 26.5%; Pred. No. 50;
Matches 36; Conservative 20; Mismatches 56; Indels 24; Gaps 7;
QY 28 GCGN--HNGGN---SSGPDYDQVTRVVTHEMAHALQSDARKSETTITQSGYNGADV 82
DB 961 GKNLVQNGGGTTLITGGNNYTG-----TTETQKTLRQGAFAFTVSSYTTIGQNGTLD 1015
QY 83 QGADNSTIELTQNGFR-----NNATIDQWNAKSDITV-GOYGGNN--AALVNTASDSS 134
DB 1016 MGGFTTISALSNNGRVLVGDNQTVGEM-----LTVAGDYSGNNGTVSLSTALADNS 1069
QY 135 VNVQVGFQGNNTANQ 150
DB 1070 KTDKLVNGSTSGTTQ 1085

RESULT 14
US-09-490-324-280
; Sequence 280, Application US/09490324
; GENERAL INFORMATION:
; APPLICANT: Knappik, Achim
; APPLICANT: Pack, Peter
; APPLICANT: Ilag, Vic
; APPLICANT: Ge, Liming
; APPLICANT: Moroney, Simon

; Plueckthun, Andreas
; TITLE OF INVENTION: Protein/(Poly)peptide libraries
; NUMBER OF SEQUENCES: 373
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Halsey, Jr., Esq. c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10021
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,324
; FILING DATE: 24-Jan-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/025,769
; FILING DATE: 18-FEB-1998
; APPLICATION NUMBER: EP 95 11 3021.0
; FILING DATE: 18-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: James F. Halsey, Jr., Esq.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)596-9090
; TELEFAX: (212)596-9090
; INFORMATION FOR SEQ ID NO: 280:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 280:
US-09-490-324-280

Query Match 10.7%; Score 83.5; DB 5; Length 415;
Best Local Similarity 29.2%; Pred. No. 12;
Matches 38; Conservative 20; Mismatches 49; Indels 23; Gaps 8;
QY 27 GGG--GNHNGGNSGPF-DYDQVLT---RVVTHEMAHALQSDAR-KSETTITQSGYNG 78
DB 250 GGSSEGGSGGSGGSGDFYKXANANKAMTENADENALQSDAKGLDSVATDYGAATD 309
QY 79 ADVQGGADNSTIELTQNGFRNNATIDQWNAKSDITVGOYGGNNALVNQ-----TASDS 133
DB 310 GFIGD-----VSLANG--NGATGDFAGSNQMAQVGE--GDNSPLMNNFRQYLPSPQ 359
QY 134 SVMVQVGFQ 143
DB 360 SVECRPFVFG 369

RESULT 15
US-09-952-267B-13
; Sequence 13, Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267B-13

Query Match      10.7%; Score 83.5; DB 5; Length 873;
Best Local Similarity 24.8%; Pred. No. 29;
Matches 40; Conservative 20; Mismatches 44; Indels 57; Gaps 10;

QY      4 LKVAAPAAIVSGSALAGVVPQWGGGNH--GGGNS-----SGPDYDQLVTRV 51
      | : | | : | | | : | | | | | | | | | | | | | | | | | |
Db      34 LLIVGILGNATTASAQQTIAQC--GKGMHSIIIGGNDNEANGDYSTVSGGDYNE----- 85

QY      52 THEMAHALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNS 111
      | : | | : | | | : | | | : | | | | | | | | | | | | | | |
Db      86 -----AKGDSSTIGGGYYNEAN-----GDSSTI---GGGFYN-----EAKGE 119

QY      112 DITVQYGGNNAALVNQRTASDSSVMRVQVFG--NNATANQY 151
      | : | | | : | | : | | : | | | | | | | | | | | | | | |
Db      120 SSTIG--GGDN-----NSATGMVSTIGGGDNNNSATCRY 150
```

Search completed: August 2, 2004, 15:29:52
Job time : 17.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec
Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| SUMMARIES | | | | | % | | Query | | Match | | Length | | DB | | ID | | Description | |
|------------|-------|-------|--------|----|--------|--|-------|--|-------|--|--------|--|----|--|----|--|---|--|
| Result No. | Score | Match | Length | DB | ID | | | | | | | | | | | | | |
| 1 | 691 | 88.6 | 151 | 2 | JC6039 | | | | | | | | | | | | fimbrin protein agfa precursor - Salmonella enteritidis | |
| 2 | 691 | 88.6 | 151 | 2 | AI0635 | | | | | | | | | | | | major curlin chain | |
| 3 | 523 | 67.1 | 151 | 2 | S70788 | | | | | | | | | | | | curlin protein cs9 | |
| 4 | 501.5 | 64.3 | 152 | 2 | D90806 | | | | | | | | | | | | curlin major subun | |
| 5 | 501.5 | 64.3 | 152 | 2 | H85665 | | | | | | | | | | | | hypothetical prote | |
| 6 | 114 | 14.6 | 1748 | 2 | S42136 | | | | | | | | | | | | cnjB protein - Tet | |
| 7 | 107.5 | 13.8 | 151 | 2 | S70787 | | | | | | | | | | | | curlin nucleator p | |
| 8 | 107.5 | 13.8 | 151 | 2 | C90806 | | | | | | | | | | | | minor curlin subun | |
| 9 | 107.5 | 13.8 | 151 | 2 | G85685 | | | | | | | | | | | | curlin minor chain | |
| 10 | 106.5 | 13.7 | 151 | 2 | JC6040 | | | | | | | | | | | | fimbrin protein ag | |
| 11 | 106.5 | 13.7 | 151 | 2 | AH0635 | | | | | | | | | | | | nucleation compone | |
| 12 | 95.5 | 12.2 | 145 | 2 | AD3143 | | | | | | | | | | | | conserved hypotet | |
| 13 | 95.5 | 12.2 | 145 | 2 | H98144 | | | | | | | | | | | | hypothetical prote | |
| 14 | 93.5 | 12.0 | 582 | 2 | F70675 | | | | | | | | | | | | probable PPE prote | |
| 15 | 93 | 11.9 | 141 | 2 | A33143 | | | | | | | | | | | | hypothetical prote | |
| 16 | 93 | 11.9 | 141 | 2 | B38145 | | | | | | | | | | | | hypothetical prote | |
| 17 | 92 | 11.8 | 401 | 2 | C88571 | | | | | | | | | | | | protein C05B5.3 [i | |
| 18 | 91.5 | 11.7 | 590 | 1 | A45621 | | | | | | | | | | | | leishmanolysin (EC | |
| 19 | 91.5 | 11.7 | 2174 | 2 | E95965 | | | | | | | | | | | | hypothetical glyci | |
| 20 | 91 | 11.7 | 590 | 2 | E70946 | | | | | | | | | | | | probable PPE prote | |
| 21 | 90 | 11.5 | 602 | 1 | PL0221 | | | | | | | | | | | | leishmanolysin (EC | |
| 22 | 90 | 11.5 | 896 | 2 | A41273 | | | | | | | | | | | | glutamate receptor | |
| 23 | 90 | 11.5 | 906 | 2 | A40222 | | | | | | | | | | | | glutamate receptor | |
| 24 | 90 | 11.5 | 906 | 2 | S25852 | | | | | | | | | | | | glutamate receptor | |
| 25 | 89.5 | 11.5 | 599 | 2 | B42049 | | | | | | | | | | | | leishmanolysin (EC | |
| 26 | 88.5 | 11.3 | 1655 | 2 | E97835 | | | | | | | | | | | | hypothetical prote | |
| 27 | 88 | 11.3 | 599 | 2 | A44951 | | | | | | | | | | | | leishmanolysin (EC | |
| 28 | 88 | 11.3 | 906 | 2 | S38723 | | | | | | | | | | | | glutamate receptor | |
| 29 | 87.5 | 11.2 | 552 | 2 | D70604 | | | | | | | | | | | | Probable PPE prote | |

DNA-binding protei
ovo protein - frui
ice nucleation pro
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical glyci
C05B5.3 protein (C
probable PPE prote
probable PPE prote
probable RTX fami
hypothetical prote
leishmanolysin (EC
hypothetical prote
chorion E2 protein
leishmanolysin (EC

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A/Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A/Reference number: JC6039; MUID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:91184712; PIDN:AAC43599.1; PID:91184714
A/Accession: PC6015
A/Molecule type: protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonell
A/Reference number: A44898; MUID:91310586; PMID:1677357

A/Contents: 27655
A/Accession: A44898
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBIP:45936)
C/Genetics:
A/Gene: agfa
C/Function:
A/Description: major component of thin aggregative fimbriae
A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbria
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 88.6%; Score 691; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 9.7e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDYDQLVTRVVTHEMAHALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYGYGSANAALALQ 60
QY 61 SPARKSETTITSGYGVNGADVCGGADNNTIELTQNGFRNNATIDOWNAKNSDITVGYQYG 120
Db 61 SPARKSETTITSGYGVNGADVCGGADNNTIELTQNGFRNNATIDOWNAKNSDITVGYQYG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2
AI0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mouton, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Mouton, S.; O'Garra, P.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0635
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 88.6%; Score 691; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 9.7e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGHNGGSGPDYDQVTRVVTHEMAHALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGHNGGSGPDYDQVTRVVTHEMAHALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVGOYGG 120
Db 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3
S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and curli
A:Reference number: S70788; MUID:96414468; PMID:8817489
A:Accession: S70788
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:gl147558; PIDN:CAR62282.1; PID:gl147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64846
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:gl1787265; PIDN:AAC74126.1; PID:gl1787279;
A:Experimental source: strain K-12, substrain MGL655
R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6,'V',8-151 <OLG1>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42;44-50 <OLS2>
R:Olsen, A.N.; Arngvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133,'RQDSGLWML' <OLS3>
A:Cross-references: EMBL:L04979; NID:g250424; PIDN:AAA23616.1; PID:g290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 523; DB 2; Length 151;
Best Local Similarity 68.9%; Pred. No. 3.1e-38;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGHNGGSGPDYDQVTRVVTHEMAHALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGHNGGSGPDYDQVTRVVTHEMAHALQ 60

QY 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVGOYGG 120
Db 61 SPARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4
D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain R1MD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA000007; PID:gl3360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs1420

Query Match 64.3%; Score 501.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 2.3e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGHNGGSGPDYDQVTRVVTHEMAHALQ 59
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGNGHNGGSGPDYDQVTRVVTHEMAHALQ 60

QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVGOYGG 119

Db 61 QADARNSDLTATQRCGGGAGDVGGSDSSIDLTLQRCFGNSATLDQWNGKDSHMTVKQFG 120
QY 120 GNNALVNOTASDSSVMVROVGFNNATANCY 151
Db 121 GGGAADVDTASNTVNTVQVGFNNATAHQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: H85665
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: H85665
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-152 <STO>
A/Cross-references: GB:AE005174; NID:gl2514574; PIDN:AAG55788.1; GSPDB:GN00145; UWSP:Z16
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: csgA

Query Match 64.3%; Score 501.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 2.3e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAFAIIVSGSALAGVPCW-CGGGNHNGGNSGPDYDQLVTRVVTHEMAHAL 59
Db 1 MKLLKVAFAIIVSGSALAGVPCW-CGGGNHNGGNSGPDYDQLVTRVVTHEMAHAL 60

QY 60 QSDARKSETTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQWAKNSDITVGGY 119
Db 61 QADARNSDLTATQRCGGGAGDVGGSDSSIDLTLQRCFGNSATLDQWNGKDSHMTVKQFG 120

QY 120 GNNALVNOTASDSSVMVROVGFNNATANCY 151
Db 121 GGGAADVDTASNTVNTVQVGFNNATAHQY 152

RESULT 6
S42136
cnjB protein - Tetrahymena thermophila
C/Species: Tetrahymena thermophila
C/Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C/Accession: S42136; S42135; S03650
R/Taylor, F.M.; Martindale, D.W.
submitted to the EMBL Data Library, October 1992
A/Reference number: S42136
A/Accession: S42136
A/Molecule type: DNA
A/Residues: 1-1748 <TAY>
A/Cross-references: EMBL:L03710; NID:gl161751; PID:gl161752
R/Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A/Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c
A/Reference number: S42135; MUID:94051569; PMID:8233798
A/Accession: S42135
A/Molecule type: DNA
A/Residues: 1164-1174, 1179-1198, 1233-1252, 1285-1293, 1316-1326, 1331-1341, 1343-1
A/Cross-references: EMBL:L03710
R/Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A/Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A/Reference number: S03650; MUID:88189811; PMID:3357771
A/Accession: S03650
A/Molecule type: DNA
A/Residues: 236-250, 'I', 252-255, 'N', 257-773 <NAR>

A/Cross-references: EMBL:X06462
C/Genetics:
A/Gene: cnjB
A/Genetic code: SGCS
A/Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C/Keywords: zinc finger
F/1164-1450/Region: glycine-rich
F/1451-1464/Region: zinc finger CCHC motif
F/1478-1491/Region: zinc finger CCHC motif
F/1501-1514/Region: zinc finger CCHC motif
F/1530-1543/Region: zinc finger CCHC motif
F/1555-1568/Region: zinc finger CCHC motif
F/1579-1592/Region: zinc finger CCHC motif
F/1602-1615/Region: zinc finger CCHC motif
F/1626-1748/Region: glycine-rich

Query Match 14.6%; Score 114; DB 2; Length 1748;
Best Local Similarity 30.5%; Pred. No. 0.11;
Matches 39; Conservative 22; Mismatches 27; Indels 40; Gaps 8;

QY 25 QWGGGHNHGG---GNSGPDYDQLVTRVVTHEMAHALQSDARKSETTTTQSGYNGADV 81
Db 1640 QFGGGGNSGGSGWGTSSGDWN-----CQSNVQES-TITSSGGWGS---- 1680

QY 82 GCGADNSTIELTQNGFRNNATIDQWAKNSDITVGGYGGNNAALVNQTSADSSVMVROVQ 141
Db 1681 -SGGNGQ---TGGWGSN---DNQOQNNENTGGGWSNS---NOTNNES----- 1722

QY 142 FGNNATAN 149
Db 1723 WGSNNQAS 1730

RESULT 7
S70787
curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
N/Alternate names: csgB protein; curlin nucleation component; minor curlin protein
C/Species: Escherichia coli
C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C/Accession: S70787; F64846
R/Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A/Title: Expression of two csg operons is required for production of fibronectin- and C
A/Reference number: S70783; MUID:96414458; PMID:8817489
A/Accession: S70787
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-151 <HAM>
A/Cross-references: EMBL:X90754; NID:gl147558; PIDN:CAA62281.1; PID:gl147563
A/Experimental source: strain K12, substrain W3110
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: F64846
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-151 <BLAT>
A/Cross-references: GB:AE000205; GB:U00096; NID:gl1787265; PIDN:AACT4125.1; PID:gl1787278
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: csgB
A/Map position: 23.15
C/Function:
A/Description: minor component of wild-type curli; interaction between CsgA and CsgB r
A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers
F/1-21/Domain: signal sequence #status predicted <SIG>
F/22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.8%; Score 107.5; DB 2; Length 151;

Best Local Similarity 35.4%; Pred. No. 0.024; Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
 Db 49 ICQAGTNNAAQLRQGGSKLLAVVAQEGSSNRKIDQTDYNL-AVIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
 Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 8
 minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subsp. C; Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: C90806
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A95629; MUID:21156231; PMID:11258796
 A;Accession: C90806
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <STO>
 A;Cross-references: GB:BA000007; PIDN:BA034842.1; PID:gl3360879; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: Ecs1419

Query Match 13.8%; Score 107.5; DB 2; Length 151;
 Best Local Similarity 35.4%; Pred. No. 0.024; Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
 Db 49 ICQAGTNNAAQLRQGGSKLLAVVAQEGSSNRKIDQTDYNL-AVIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
 Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 9
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7; Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: G85665
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: G85665
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <STO>
 A;Cross-references: GB:AB005174; NID:gl2514573; PIDN:AA055787.1; GSPDB:GN00145; UWGP:216
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: csgB

Query Match 13.8%; Score 107.5; DB 2; Length 151;
 Best Local Similarity 35.4%; Pred. No. 0.024; Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
 Db 49 ICQAGTNNAAQLRQGGSKLLAVVAQEGSSNRKIDQTDYNL-AVIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
 Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 10
 fimbrin protein agfB precursor - Salmonella enteritidis
 C;Species: Salmonella enteritidis
 C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
 C;Accession: JC6040
 R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W. J. Bacteriol. 178, 662-667, 1996
 A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae. A;Reference number: JC6039; MUID:96146512; PMID:8550497
 A;Accession: JC6040
 A;Molecule type: DNA
 A;Residues: 1-151 <COL>
 A;Cross-references: GB:U43280; NID:gl184712; PIDN:AA043598.1; PID:gl184713
 A;Experimental source: strain 276755-3b
 C;Genetics:
 A;Gene: agfB
 C;Function:
 A;Description: minor component of thin aggregative fimbrinae
 A;Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator
 C;Keywords: fimbrina
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-151/Product: fimbrin protein agfB #status predicted <MAT>

Query Match 13.7%; Score 106.5; DB 2; Length 151;
 Best Local Similarity 30.7%; Pred. No. 0.029; Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQGGADNST-----IELTQNGFR 98
 Db 18 IATATNYDLARSEYNFAVNELSKSFNQAAIIQVGTDSARVRQEGSKLLSVISQEGN 77

QY 99 NNATIDQWNAKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 Db 78 NRAKVDQ--AGNYNFAIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 11
 AH0635
 nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica C;Species: Salmonella enterica subsp. enterica serovar Typhi
 A;Note: this species has also been called Salmonella typhi
 C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C;Accession: AH0635
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, Th. T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K. A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi. A;Reference number: AB0502; MUID:21534947; PMID:11677608
 A;Accession: AH0635
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <PAR>
 A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl650233.4; GSPDB:GN00176
 C;Genetics:
 A;Gene: STY1180

Query Match 13.7%; Score 106.5; DB 2; Length 151;
 Best Local Similarity 30.7%; Pred. No. 0.029; Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQGGADNST-----IELTQNGFR 98
 Db 18 IATATNYDLARSEYNFAVNELSKSFNQAAIIQVGTDSARVRQEGSKLLSVISQEGN 77

QY 99 NNATIDQWNAKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 750

Sequence: 1 MKLLKVAAPFAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 691 | 88.6 | 151 | 1 | CSGA_SALTY |
| 2 | 523 | 67.1 | 151 | 1 | CSGA_ECOLI |
| 3 | 501.5 | 64.3 | 152 | 1 | CSGA_ECO57 |
| 4 | 107.5 | 13.8 | 151 | 1 | CSGB_ECOLI |
| 5 | 106.5 | 13.7 | 151 | 1 | CSGB_SALTY |
| 6 | 106.5 | 13.7 | 151 | 1 | CSGB_SALTY |
| 7 | 92 | 11.8 | 401 | 1 | YK03_CAREL |
| 8 | 92 | 11.8 | 1656 | 1 | OMP8_RICGA |
| 9 | 91.5 | 11.7 | 590 | 1 | GP63_LEIDO |
| 10 | 91 | 11.7 | 1093 | 1 | PER_DROWI |
| 11 | 90 | 11.5 | 602 | 1 | GP63_LEIMA |
| 12 | 90 | 11.5 | 906 | 1 | GUR1_HUMAN |
| 13 | 88.5 | 11.3 | 1655 | 1 | OMP8_RICON |
| 14 | 88 | 11.3 | 599 | 1 | GP63_LEICH |
| 15 | 87.5 | 11.2 | 1028 | 1 | OVN_DROME |
| 16 | 87.5 | 11.2 | 1567 | 1 | ICEN_XANCT |
| 17 | 86 | 11.0 | 485 | 1 | Y136_TREPA |
| 18 | 86 | 11.0 | 493 | 1 | GATA_RHIME |
| 19 | 84 | 10.8 | 646 | 1 | GP63_LEIME |
| 20 | 84 | 10.8 | 907 | 1 | GUR1_MOUSE |
| 21 | 84 | 10.8 | 907 | 1 | GUR1_RAT |
| 22 | 83 | 10.6 | 720 | 1 | KKE6_YEAST |
| 23 | 83 | 10.6 | 947 | 1 | SECA_STRCO |
| 24 | 82 | 10.5 | 342 | 1 | OMP8_RAHQA |
| 25 | 81.5 | 10.4 | 424 | 1 | COAA_BPFD |
| 26 | 81.5 | 10.4 | 424 | 1 | COAA_BPFI3 |
| 27 | 81.5 | 10.4 | 493 | 1 | GATA_AGRF5 |
| 28 | 81 | 10.4 | 365 | 1 | ROAL_DROME |
| 29 | 81 | 10.4 | 576 | 1 | DEAF_DROME |
| 30 | 81 | 10.4 | 678 | 1 | YF48_MYCTU |
| 31 | 81 | 10.4 | 1034 | 1 | ICEN_PANAN |
| 32 | 80.5 | 10.3 | 641 | 1 | IND_ARTGO |
| 33 | 80.5 | 10.3 | 1185 | 1 | MAPX_DROME |

34 80 10.3 347 1 MSA2_PLAF2
35 79.5 10.2 940 1 SECA_STRGR
36 79.5 10.2 1300 1 L2PK_RICRI
37 79.5 10.2 1654 1 OMP8_RICRI
38 79 10.1 199 1 AC22_TENMO
39 79 10.1 1140 1 YM96_YEAST
40 78.5 10.1 704 1 LIEM_NEIMA
41 78.5 10.1 1293 1 MLE_DROME
42 78 10.0 485 1 YB47_MYCPN
43 78 10.0 588 1 CAR7_CANAL
44 78 10.0 1258 1 ICEN_ERWHE
45 78 10.0 2003 1 YDBA_ECOLI

Q03646 plasmodium
P95759 streptomyce
P14914 rickettsia
Q53047 r outer mem
P26968 tenebrio mo
Q04893 saccharomyc
P57037 neisseria m
P24785 drosophila
P75139 mycoplasma
P43096 candida alb
P16239 erwinia her
P33666 escherichia

ALIGNMENTS

RESULT 1
CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGPA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1] Sequence from N.A.
RP SPECIES=Salmonella typhimurium; STRAIN=SR-11;
RC MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";
RL Nature 413:852-856 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Salmonella typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

```

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL and Ctl8.";
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RL J. Bacteriol. 178:662-667(1996).
RN [6]
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae.";
RL J. Clin. Microbiol. 31:2263-2273(1993).
RN [7]
RP SEQUENCE OF 21-33.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeoey L., Mueller K.-M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ002301; CAA05317.1; -
DR EMBL; AF008749; AAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; AF016840; AAO69399.1; -
DR EMBL; U43280; AAC43599.1; -
DR PIR; JC6039; JC6039.
DR StyGene; SG10608; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SYMVRQVGFNNATANYQ -> DSYTQVAS (IN
REF. 6).
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
Query Match 88.6%; Score 691; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 5.3e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGNPGDPYQLVTRVVTHEMAHALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGNPGDSTLSIVQYGSNAALALQ 60
QY 61 SPARKSETTIQSGYNGADVCGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Dd 61 SPARKSETTIQSGYNGADVCGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVWVQVGFNNATANYQ 151
Dd 121 NNAALVNOTASDSSVWVQVGFNNATANYQ 151

```

```

RESULT 2
CSGA_ECOLI
ID CSGA_ECOLI STANDARD; PRT; 151 AA.
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / WJ110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli";
RT Escherichia coli;
RL Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
fibronectin- and congo red-binding curli polymers in Escherichia coli
K-12.";
RT Mol. Microbiol. 18:661-670(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itch T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
Sampai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YMEJ;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
fibronectin binding in Escherichia coli HB101.";
RL Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeoey L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RL J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRONECTIN.
CC

```

```

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04579; AAA23616.1; -
CC DR EMBL; X90754; CAA62282.1; -
CC DR EMBL; AAC00205; AAC74126.1; -
CC DR EMBL; D90741; BAA35832.1; -
CC DR EMBL; D90742; BAA35840.1; -
CC DR PIR; S70788; S70788.
CC ECoGene; BG11489; CSGA.
CC Fimbrria; Signal; Complete proteome.
KW SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 A -> E (IN REF. 1).
FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
SQ
Query Match 67.1%; Score 523; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.7e-38;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQVGCGGNHGGNSGSDPDYDQLVTRVVTHEMAHALQ 60
Db 1 MKLLKVAIAAIVFSGSALAGVVPQVGCGGNHGGNSGNSPSEINITYQGGNSALALQ 60
Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNATIDQWNAKNSDITVGYQG 120
Db 61 TDARNSDLTITQHGCGNGADVGCGSDSDSIDLTQSGFGNSATLDQWNGKNSMTVTKQFG 120
Qy 121 NNAALVNQATSDSSVYVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSSVYVTVQVGFNNATANQY 151
RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Uhlrich G.A., Keen J.E., Elder R.O.,
RT "Mutations in the csgr promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RT Nature 409:529-533 (2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Ohnishi M., Kurata K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE275733; AAK53212.1; -
CC DR EMBL; AE005315; ARG55788.1; -
CC DR EMBL; AP002554; BAB34843.1; -
CC DR PIR; D90806; D90806.
CC DR PIR; H85665; H85665.
CC Fimbrria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
SQ
Query Match 64.3%; Score 501.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 1.2e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQVGCGGNHGGNSGSDPDYDQLVTRVVTHEMAHALQ 59
Db 1 MKLLKVAIAAIVFSGSALAGVVPQVGCGGNHGGNSGNSPSEINITYQGGNSALAL 60
Qy 60 QSDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNATIDQWNAKNSDITVGYQG 119
Db 61 QADARNSDLTITQHGCGNGADVGCGSDSDSIDLTQSGFGNSATLDQWNGKNSMTVTKQFG 120
Qy 120 GNNALVNQATSDSSVYVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSSVYVTVQVGFNNATANQY 152
RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR E1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arrigvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csgr operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli

```

```

RT RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:11433-1474 (1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.F., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi K., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [6]
RN SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arqvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csxB promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
of the nucleoid-associated protein H-NS."
RL Mol. Microbiol. 13:1021-1032(1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/)
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; X90754; CAA62281.1; -.

```

```

DR EMBL; AE000205; AAC74125.1; -.
DR EMBL; D90741; BAA35831.1; -.
DR EMBL; AE005315; AAG55787.1; -.
DR EMBL; AP002554; BAB34842.1; -.
DR FIR; C90806; C90806.
DR FIR; G85665; G85665.
DR FIR; S70787; S70787.
DR EcoGene; EG12621; csxB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match 13.8%; Score 107.5; DB 1; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.012;
Matches 29; Conservative

QY 70 ITQSGYNGADVGGADNNTIELTONGFRNATIDWNKNSDITVGYQGNRAALVNQT 129
DB 49 IQQAGTNNSQLRQGGSKLLAVVAQEGSSNRKIDQTGYNL-AYIDQASANDASISQ 107
QY 130 ASDSSVMVQVGFNNATANOY 151
DB 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 5
CSGB SALT1 STANDARD; PRT; 151 AA.
ID _CSGB_SALT1 STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kocoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AL627289; CAD08287.1; -;
CC DR EMBL; AE016840; AAO69400.1; -;
CC KW Fimbria; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 13.7%; Score 106.5; DB 1; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.015;
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQ--GADNST-----IELTQNGFR 98
DB 18 IATATNYDLARSEYFVAVNELSKSFNCAAIIGVGTDSARVQRQSGKLLSVISQEGN 77
QY 99 NNATIDQWNAKSDIT-VQYGGNNALVNOTASDSSVMVRQVGFNNATANQY 151
DB 78 NRAKVQDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 6
CSGB SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC11412 / ATCC 700720;
RX MEDLINE=2153498; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouchier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae";
RL J. Bacteriol. 178:662-667(1996).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AJ002301; CAA05316.1; -;
CC DR EMBL; AB008749; AAL20073.1; -;
CC DR EMBL; U43280; AAC43598.1; -;
CC DR PIR; JC6040; JC6040.
CC DR StyGene; SG10609; CSGB.
CC KW Fimbria; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;

Query Match 13.7%; Score 106.5; DB 1; Length 151;
Best Local Similarity 30.7%; Pred. No. 0.015;
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQ--GADNST-----IELTQNGFR 98
DB 18 IATATNYDLARSEYFVAVNELSKSFNCAAIIGVGTDSARVQRQSGKLLSVISQEGN 77
QY 99 NNATIDQWNAKSDIT-VQYGGNNALVNOTASDSSVMVRQVGFNNATANQY 151
DB 78 NRAKVQDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

RESULT 7
YK03 CAEL STANDARD; PRT; 401 AA.
ID -YK03 CAEL
AC P34291;
DT 01-FEB-1994 (Rel. 28; Created)
DT 30-MAY-2000 (Rel. 39; Last sequence update)
DT 30-MAY-2000 (Rel. 39; Last annotation update)
DE Hypothetical 43.5 kDa protein C05B5.3 in chromosome III.
GN C05B5.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Paloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Mortimore B.J.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

CC EMBL; Z32679; CAA83596.1; -;
CC DR PIR; C88571; C88571.
CC DR WormPep; C05B5.3; CE17369.
CC KW Hypothetical protein.
CC SQ SEQUENCE 401 AA; 43472 MW; C99DF259EF6C8B55 CRC64;

Query Match 11.8%; Score 92; DB 1; Length 401;
Best Local Similarity 23.2%; Pred. No. 0.8;
Matches 46; Conservative 19; Mismatches 71; Indels 62; Gaps 7;

```

QY 3 LLKVAFAAIVSGSALAGVVPQWGG-----GNHNGGNSGPDYDQLVT 48
Db 30 LQYRAYAGSGSGCPATIVPKSGFWPNADMIAGLQTEQSRQNSQNNPQQDDPT 89
QY 49 RVVTHE-----MAHALQSDARKSETTI-----TQSG--YGNADVGGQA 85
Db 90 SQSTGOINGNVGSSSSNQPVIIYIARAGSKYKXSEVTSTPTNGFNGFGQGNQ 149
QY 86 D-----NSTIETQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOQTASD 132
Db 150 NTGFSGFFNNQNSQNLNQNFOQN-----QNLGASGFFNNQN-----QNSQQ 197
QY 133 SSVNVRQVGFNNATNQ 150
Db 198 NVNGPTSGFSNQTSNQ 215

RESULT 8
OMP_B_RICJA STANDARD; PRT; 1656 AA.
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rompB)
DE (romp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=357930;
RN SEQUENCE FROM N.A.
RP STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -
CC InterPro; IPR006315; Autotransporter.
CC InterPro; IPR005546; Autotransporter.
CC Pfam; PF03797; Autotransporter; 1.
CC TIGRFAMs; TIGR01414; autotrans_barl; 2.
CC Antigen; S-layer; Cell wall.
CC CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
CC FT DOMAIN 528 533 POLY-GLY.
CC SEQUENCE 1656 AA; 168097 MM; 3132A69C9DD5999F CRC64;
SQ
Query Match 11.8%; Score 92; DB 1; Length 1656;
Best Local Similarity 26.2%; Pred. No. 4;
Matches 45; Conservative 19; Mismatches 52; Indels 56; Gaps 9;

```

```

QY 6 VAAFAAIVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTVVTHEMAHALQSDARK 65
Db 509 VLAGAITLQGSATI-----TGDIQGGGAA-----LQSI-----LANDATK 547
QY 66 SETTITQSG-----YGNADVGGQADNSTIETQNGFRNNATID----- 104
Db 548 ---TLTGGAGIISANGGTINFQANGGTIKLTST--QNNIIVDCDLAIAIDQGTGVVDASS 602
QY 105 QWNAKNSDI--TVGOYGGNNAAL-----VNQTSDDSSVWVQVGFNNAT 147
Db 603 LTNAQTITISGTITIGANNNTTIGQFNGSKTTLGNGVAINELVIGNNGS 654

RESULT 9
GP63 LEIDO STANDARD; PRT; 590 AA.
AC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RA MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
RT of Leishmania donovani."
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr--Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M60048; AAA29244.1; -
CC HSP; P08148; LMML.
CC DR MEROPS; M08.001; -
CC InterPro; IPR006025; Pept M Zn BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase_M8; 1.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00442; ZINC_PROTEASE; 1.
CC Hydrolyase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
CC SIGNAL 1 39 POTENTIAL.
CC PROPEP 40 87 ACTIVATION PEPTIDE.
CC CHAIN 88 565 LEISHMANOLYSIN.
CC FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
CC FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 252 252 BY SIMILARITY.
CC FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DISULFID 112 129 BY SIMILARITY.
CC FT DISULFID 178 217 BY SIMILARITY.
CC FT DISULFID 301 373 BY SIMILARITY.

```


FT DISULFID 380 443 BY SIMILARITY.
 FT DISULFID 393 412 BY SIMILARITY.
 FT DISULFID 402 477 BY SIMILARITY.
 FT DISULFID 454 498 BY SIMILARITY.
 FT DISULFID 503 553 BY SIMILARITY.
 FT DISULFID 523 546 BY SIMILARITY.
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299595F58 CRC64;
 Query Match 11.7%; Score 91.5; DB 1; Length 590;
 Best Local Similarity 30.5%; Pred. No. 1.4;
 Matches 43; Conservative 12; Mismatches 39; Indels 47; Gaps 9;
 QY 43 YDQLVTRVVTHEMAHALQ-----SDAR-----KSETTITOS--GYG 76
 Db 241 YDQLVTRVVTHEMAHALGFSVVFFRDARILESINVRHKDFDPVINSSTAVAKRQYG 300
 QY 77 NGA-----DVGQGANSTTELQNGFRNNATTQDNNAKNSDITVQYGGNNAALVQ 128
 Db 301 CGTLEYLEMEDQGGAGSAGSHIKM-----RNAQ-DELMAPASD--AGYYSALTMAIFQD 351
 QY 129 TA---SDSSVMVRQVGFQGNNA 146
 Db 352 LCFYQADPS-KAEEMPWGRNA 371

RESULT 10

PER DROWI
 ID PER DROWI STANDARD; PRT; 1093 AA.
 AC Q03297; O18421; O18422; P91721; P91722;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Period circadian protein (Fragment).
 GN PER.
 OS Drosophila willistoni (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7260;
 RN [1]
 RN [2]
 SEQUENCE FROM N.A.
 RC STRAIN=Various strains;
 RX MEDLINE=97357421; PubMed=8450754;
 RA Gleason J.M., Powell J.R.;
 RT "Interspecific and intraspecific comparisons of the period locus in
 the Drosophila willistoni sibling species.";
 RL Mol. Biol. Evol. 14:741-753 (1997).
 RN [1]
 RN [2]

CC -1- FUNCTION: Essential for biological clock functions. Determines the
 period length of circadian and ultradian rhythms; an increase in
 PER dosage leads to shortened circadian rhythms and a decrease
 leads to lengthened circadian rhythms. Essential for the circadian
 rhythmicity of locomotor activity, eclosion behavior, and for the
 rhythmic component of the male courtship song that originates in
 the thoracic nervous system. The biological cycle depends on the
 rhythmic formation and nuclear localization of the TIM-PER
 complex. Light induces the degradation of TIM, which promotes
 elimination of PER. Nuclear activity of the heterodimer
 coordinatively regulates PER and TIM transcription through a
 negative feedback loop. Behaves as a negative element in circadian
 transcriptional loop. Does not appear to bind DNA, suggesting
 indirect transcriptional inhibition (By similarity).
 CC -1- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
 translocates into the nucleus (By similarity).

CC -1- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
 CC First accumulates in the perinuclear region about one hour before
 CC translocation into the nucleus. Interaction with TIM is required
 CC for nuclear localization (By similarity).
 CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
 CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
 CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
 CC PER-TIM (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.
 CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; U51055; AAB41360.1; -
 CC EMBL; U51056; AAB41361.1; -
 CC EMBL; U51057; AAB41362.1; -
 CC EMBL; U51058; AAB41363.1; -
 CC EMBL; U51059; AAB41364.1; -
 CC EMBL; U51060; AAB41365.1; -
 CC EMBL; U51061; AAB41366.1; -
 CC EMBL; U51062; AAB41367.1; -
 CC EMBL; U51063; AAB41368.1; -
 CC EMBL; U51064; AAB41369.1; -
 CC EMBL; U51065; AAB41370.1; -
 CC EMBL; U51066; AAB41371.1; -
 CC EMBL; U51067; AAB41372.1; -
 CC EMBL; U51068; AAB41373.1; -
 CC EMBL; U51069; AAB41374.1; -
 CC EMBL; U51070; AAB41375.1; -
 CC EMBL; U51071; AAB41376.1; -
 CC EMBL; U51072; AAB41377.1; -
 CC EMBL; U51073; AAB41378.1; -
 CC EMBL; U51074; AAB41379.1; -
 CC EMBL; U51075; AAB41380.1; -
 CC EMBL; U51076; AAB41381.1; -
 CC EMBL; U51077; AAB41382.1; -
 CC EMBL; U51078; AAB41383.1; -
 CC EMBL; U51079; AAB41384.1; -
 CC EMBL; U51080; AAB41385.1; -
 CC EMBL; U51081; AAB41386.1; -
 CC EMBL; U51082; AAB41387.1; -
 CC EMBL; U51083; AAB41388.1; -
 CC EMBL; U51084; AAB41389.1; -
 CC EMBL; U51085; AAB41390.1; -
 CC EMBL; U51086; AAB41391.1; -
 CC EMBL; U51087; AAB41392.1; -
 CC EMBL; U51088; AAB41393.1; -
 CC EMBL; U51089; AAB41394.1; -
 CC EMBL; U51090; AAB41395.1; -
 CC EMBL; U51091; AAB41396.1; -
 CC EMBL; U51092; AAB41397.1; -
 CC EMBL; U51093; AAB41398.1; -
 CC EMBL; U51094; AAB41399.1; -
 CC EMBL; U51095; AAB41400.1; -
 CC EMBL; U51096; AAB41401.1; -
 CC EMBL; U51097; AAB41402.1; -
 CC EMBL; U51098; AAB41403.1; -
 CC EMBL; U51099; AAB41404.1; -
 CC EMBL; U51100; AAB41405.1; -
 CC EMBL; U51101; AAB41406.1; -
 CC EMBL; U51102; AAB41407.1; -
 CC EMBL; U51103; AAB41408.1; -
 CC EMBL; U51104; AAB41409.1; -
 CC EMBL; U51105; AAB41410.1; -
 CC EMBL; U51106; AAB41411.1; -
 CC EMBL; U51107; AAB41412.1; -
 CC EMBL; U51108; AAB41413.1; -
 CC EMBL; U51109; AAB41414.1; -
 CC EMBL; U51110; AAB41415.1; -
 CC EMBL; U51111; AAB41416.1; -
 CC EMBL; U51112; AAB41417.1; -
 CC EMBL; U51113; AAB41418.1; -
 CC EMBL; U51114; AAB41419.1; -
 CC EMBL; U51115; AAB41420.1; -
 CC EMBL; U51116; AAB41421.1; -
 CC EMBL; U51117; AAB41422.1; -
 CC EMBL; U51118; AAB41423.1; -
 CC EMBL; U51119; AAB41424.1; -
 CC EMBL; U51120; AAB41425.1; -
 CC EMBL; U51121; AAB41426.1; -
 CC EMBL; U51122; AAB41427.1; -
 CC EMBL; U51123; AAB41428.1; -
 CC EMBL; U51124; AAB41429.1; -
 CC EMBL; U51125; AAB41430.1; -
 CC EMBL; U51126; AAB41431.1; -
 CC EMBL; U51127; AAB41432.1; -
 CC EMBL; U51128; AAB41433.1; -
 CC EMBL; U51129; AAB41434.1; -
 CC EMBL; U51130; AAB41435.1; -
 CC EMBL; U51131; AAB41436.1; -
 CC EMBL; U51132; AAB41437.1; -
 CC EMBL; U51133; AAB41438.1; -
 CC EMBL; U51134; AAB41439.1; -
 CC EMBL; U51135; AAB41440.1; -
 CC EMBL; U51136; AAB41441.1; -
 CC EMBL; U51137; AAB41442.1; -
 CC EMBL; U51138; AAB41443.1; -
 CC EMBL; U51139; AAB41444.1; -
 CC EMBL; U51140; AAB41445.1; -
 CC EMBL; U51141; AAB41446.1; -
 CC EMBL; U51142; AAB41447.1; -
 CC EMBL; U51143; AAB41448.1; -
 CC EMBL; U51144; AAB41449.1; -
 CC EMBL; U51145; AAB41450.1; -
 CC EMBL; U51146; AAB41451.1; -
 CC EMBL; U51147; AAB41452.1; -
 CC EMBL; U51148; AAB41453.1; -
 CC EMBL; U51149; AAB41454.1; -
 CC EMBL; U51150; AAB41455.1; -
 CC EMBL; U51151; AAB41456.1; -
 CC EMBL; U51152; AAB41457.1; -
 CC EMBL; U51153; AAB41458.1; -
 CC EMBL; U51154; AAB41459.1; -
 CC EMBL; U51155; AAB41460.1; -
 CC EMBL; U51156; AAB41461.1; -
 CC EMBL; U51157; AAB41462.1; -
 CC EMBL; U51158; AAB41463.1; -
 CC EMBL; U51159; AAB41464.1; -
 CC EMBL; U51160; AAB41465.1; -
 CC EMBL; U51161; AAB41466.1; -
 CC EMBL; U51162; AAB41467.1; -
 CC EMBL; U51163; AAB41468.1; -
 CC EMBL; U51164; AAB41469.1; -
 CC EMBL; U51165; AAB41470.1; -
 CC EMBL; U51166; AAB41471.1; -
 CC EMBL; U51167; AAB41472.1; -
 CC EMBL; U51168; AAB41473.1; -
 CC EMBL; U51169; AAB41474.1; -
 CC EMBL; U51170; AAB41475.1; -
 CC EMBL; U51171; AAB41476.1; -
 CC EMBL; U51172; AAB41477.1; -
 CC EMBL; U51173; AAB41478.1; -
 CC EMBL; U51174; AAB41479.1; -
 CC EMBL; U51175; AAB41480.1; -
 CC EMBL; U51176; AAB41481.1; -
 CC EMBL; U51177; AAB41482.1; -
 CC EMBL; U51178; AAB41483.1; -
 CC EMBL; U51179; AAB41484.1; -
 CC EMBL; U51180; AAB41485.1; -
 CC EMBL; U51181; AAB41486.1; -
 CC EMBL; U51182; AAB41487.1; -
 CC EMBL; U51183; AAB41488.1; -
 CC EMBL; U51184; AAB41489.1; -
 CC EMBL; U51185; AAB41490.1; -
 CC EMBL; U51186; AAB41491.1; -
 CC EMBL; U51187; AAB41492.1; -
 CC EMBL; U51188; AAB41493.1; -
 CC EMBL; U51189; AAB41494.1; -
 CC EMBL; U51190; AAB41495.1; -
 CC EMBL; U51191; AAB41496.1; -
 CC EMBL; U51192; AAB41497.1; -
 CC EMBL; U51193; AAB41498.1; -
 CC EMBL; U51194; AAB41499.1; -
 CC EMBL; U51195; AAB41500.1; -
 CC EMBL; U51196; AAB41501.1; -
 CC EMBL; U51197; AAB41502.1; -
 CC EMBL; U51198; AAB41503.1; -
 CC EMBL; U51199; AAB41504.1; -
 CC EMBL; U51200; AAB41505.1; -
 CC EMBL; U51201; AAB41506.1; -
 CC EMBL; U51202; AAB41507.1; -
 CC EMBL; U51203; AAB41508.1; -
 CC EMBL; U51204; AAB41509.1; -
 CC EMBL; U51205; AAB41510.1; -
 CC EMBL; U51206; AAB41511.1; -
 CC EMBL; U51207; AAB41512.1; -
 CC EMBL; U51208; AAB41513.1; -
 CC EMBL; U51209; AAB41514.1; -
 CC EMBL; U51210; AAB41515.1; -
 CC EMBL; U51211; AAB41516.1; -
 CC EMBL; U51212; AAB41517.1; -
 CC EMBL; U51213; AAB41518.1; -
 CC EMBL; U51214; AAB41519.1; -
 CC EMBL; U51215; AAB41520.1; -
 CC EMBL; U51216; AAB41521.1; -
 CC EMBL; U51217; AAB41522.1; -
 CC EMBL; U51218; AAB41523.1; -
 CC EMBL; U51219; AAB41524.1; -
 CC EMBL; U51220; AAB41525.1; -
 CC EMBL; U51221; AAB41526.1; -
 CC EMBL; U51222; AAB41527.1; -
 CC EMBL; U51223; AAB41528.1; -
 CC EMBL; U51224; AAB41529.1; -
 CC EMBL; U51225; AAB41530.1; -
 CC EMBL; U51226; AAB41531.1; -
 CC EMBL; U51227; AAB41532.1; -
 CC EMBL; U51228; AAB41533.1; -
 CC EMBL; U51229; AAB41534.1; -
 CC EMBL; U51230; AAB41535.1; -
 CC EMBL; U51231; AAB41536.1; -
 CC EMBL; U51232; AAB41537.1; -
 CC EMBL; U51233; AAB41538.1; -
 CC EMBL; U51234; AAB41539.1; -
 CC EMBL; U51235; AAB41540.1; -
 CC EMBL; U51236; AAB41541.1; -
 CC EMBL; U51237; AAB41542.1; -
 CC EMBL; U51238; AAB41543.1; -
 CC EMBL; U51239; AAB41544.1; -
 CC EMBL; U51240; AAB41545.1; -
 CC EMBL; U51241; AAB41546.1; -
 CC EMBL; U51242; AAB41547.1; -
 CC EMBL; U51243; AAB41548.1; -
 CC EMBL; U51244; AAB41549.1; -
 CC EMBL; U51245; AAB41550.1; -
 CC EMBL; U51246; AAB41551.1; -
 CC EMBL; U51247; AAB41552.1; -
 CC EMBL; U51248; AAB41553.1; -
 CC EMBL; U51249; AAB41554.1; -
 CC EMBL; U51250; AAB41555.1; -
 CC EMBL; U51251; AAB41556.1; -
 CC EMBL; U51252; AAB41557.1; -
 CC EMBL; U51253; AAB41558.1; -
 CC EMBL; U51254; AAB41559.1; -
 CC EMBL; U51255; AAB41560.1; -
 CC EMBL; U51256; AAB41561.1; -
 CC EMBL; U51257; AAB41562.1; -
 CC EMBL; U51258; AAB41563.1; -
 CC EMBL; U51259; AAB41564.1; -
 CC EMBL; U51260; AAB41565.1; -
 CC EMBL; U51261; AAB41566.1; -
 CC EMBL; U51262; AAB41567.1; -
 CC EMBL; U51263; AAB41568.1; -
 CC EMBL; U51264; AAB41569.1; -
 CC EMBL; U51265; AAB41570.1; -
 CC EMBL; U51266; AAB41571.1; -
 CC EMBL; U51267; AAB41572.1; -
 CC EMBL; U51268; AAB41573.1; -
 CC EMBL; U51269; AAB41574.1; -
 CC EMBL; U51270; AAB41575.1; -
 CC EMBL; U51271; AAB41576.1; -
 CC EMBL; U51272; AAB41577.1; -
 CC EMBL; U51273; AAB41578.1; -
 CC EMBL; U51274; AAB41579.1; -
 CC EMBL; U51275; AAB41580.1; -
 CC EMBL; U51276; AAB41581.1; -
 CC EMBL; U51277; AAB41582.1; -
 CC EMBL; U51278; AAB41583.1; -
 CC EMBL; U51279; AAB41584.1; -
 CC EMBL; U51280; AAB41585.1; -
 CC EMBL; U51281; AAB41586.1; -
 CC EMBL; U51282; AAB41587.1; -
 CC EMBL; U51283; AAB41588.1; -
 CC EMBL; U51284; AAB41589.1; -
 CC EMBL; U51285; AAB41590.1; -
 CC EMBL; U51286; AAB41591.1; -
 CC EMBL; U51287; AAB41592.1; -
 CC EMBL; U51288; AAB41593.1; -
 CC EMBL; U51289; AAB41594.1; -
 CC EMBL; U51290; AAB41595.1; -
 CC EMBL; U51291; AAB41596.1; -
 CC EMBL; U51292; AAB41597.1; -
 CC EMBL; U51293; AAB41598.1; -
 CC EMBL; U51294; AAB41599.1; -
 CC EMBL; U51295; AAB41600.1; -
 CC EMBL; U51296; AAB41601.1; -
 CC EMBL; U51297; AAB41602.1; -
 CC EMBL; U51298; AAB41603.1; -
 CC EMBL; U51299; AAB41604.1; -
 CC EMBL; U51300; AAB41605.1; -
 CC EMBL; U51301; AAB41606.1; -
 CC EMBL; U51302; AAB41607.1; -
 CC EMBL; U51303; AAB41608.1; -
 CC EMBL; U51304; AAB41609.1; -
 CC EMBL; U51305; AAB41610.1; -
 CC EMBL; U51306; AAB41611.1; -
 CC EMBL; U51307; AAB41612.1; -
 CC EMBL; U51308; AAB41613.1; -
 CC EMBL; U51309; AAB41614.1; -
 CC EMBL; U51310; AAB41615.1; -
 CC EMBL; U51311; AAB41616.1; -
 CC EMBL; U51312; AAB41617.1; -
 CC EMBL; U51313; AAB41618.1; -
 CC EMBL; U51314; AAB41619.1; -
 CC EMBL; U51315; AAB41620.1; -
 CC EMBL; U51316; AAB41621.1; -
 CC EMBL; U51317; AAB41622.1; -
 CC EMBL; U51318; AAB41623.1; -
 CC EMBL; U51319; AAB41624.1; -
 CC EMBL; U51320; AAB41625.1; -
 CC EMBL; U51321; AAB41626.1; -
 CC EMBL; U51322; AAB41627.1; -
 CC EMBL; U51323; AAB41628.1; -
 CC EMBL; U51324; AAB41629.1; -
 CC EMBL; U51325; AAB41630.1; -
 CC EMBL; U51326; AAB41631.1; -
 CC EMBL; U51327; AAB41632.1; -
 CC EMBL; U51328; AAB41633.1; -
 CC EMBL; U51329; AAB41634.1; -
 CC EMBL; U51330; AAB41635.1; -
 CC EMBL; U51331; AAB41636.1; -
 CC EMBL; U51332; AAB41637.1; -
 CC EMBL; U51333; AAB41638.1; -
 CC EMBL; U51334; AAB41639.1; -
 CC EMBL; U51335; AAB41640.1; -
 CC EMBL; U51336; AAB41641.1; -
 CC EMBL; U51337; AAB41642.1; -
 CC EMBL; U51338; AAB41643.1; -
 CC EMBL; U51339; AAB41644.1; -
 CC EMBL; U51340; AAB41645.1; -
 CC EMBL; U51341; AAB41646.1; -
 CC EMBL; U51342; AAB41647.1; -
 CC EMBL; U51343; AAB41648.1; -
 CC EMBL; U51344; AAB41649.1; -
 CC EMBL; U51345; AAB41650.1; -
 CC EMBL; U51346; AAB41651.1; -
 CC EMBL; U51347; AAB41652.1; -
 CC EMBL; U51348; AAB41653.1; -
 CC EMBL; U51349; AAB41654.1; -
 CC EMBL; U51350; AAB41655.1; -
 CC EMBL; U51351; AAB41656.1; -
 CC EMBL; U51352; AAB41657.1; -
 CC EMBL; U51353; AAB41658.1; -
 CC EMBL; U51354; AAB41659.1; -
 CC EMBL; U51355; AAB41660.1; -
 CC EMBL; U51356; AAB41661.1; -
 CC EMBL; U51357; AAB41662.1; -
 CC EMBL; U51358; AAB41663.1; -
 CC EMBL; U51359; AAB41664.1; -
 CC EMBL; U51360; AAB41665.1; -
 CC EMBL; U51361; AAB41666.1; -
 CC EMBL; U51362; AAB41667.1; -
 CC EMBL; U51363; AAB41668.1; -
 CC EMBL; U51364; AAB41669.1; -
 CC EMBL; U51365; AAB41670.1; -
 CC EMBL; U51366; AAB41671.1; -
 CC EMBL; U51367; AAB41672.1; -
 CC EMBL; U51368; AAB41673.1; -
 CC EMBL; U51369; AAB41674.1; -
 CC EMBL; U51370; AAB41675.1; -
 CC EMBL; U51371; AAB41676.1; -
 CC EMBL; U51372; AAB41677.1; -
 CC EMBL; U51373; AAB41678.1; -
 CC EMBL; U51374; AAB41679.1; -
 CC EMBL; U51375; AAB41680.1; -
 CC EMBL; U51376; AAB41681.1; -
 CC EMBL; U51377; AAB41682.1; -
 CC EMBL; U51378; AAB41683.1; -
 CC EMBL; U51379; AAB41684.1; -
 CC EMBL; U51380; AAB41685.1; -
 CC EMBL; U51381; AAB41686.1; -
 CC EMBL; U51382; AAB41687.1; -
 CC EMBL; U51383; AAB41688.1; -
 CC EMBL; U51384; AAB41689.1; -
 CC EMBL; U51385; AAB41690.1; -
 CC EMBL; U51386; AAB41691.1; -
 CC EMBL; U51387; AAB41692.1; -
 CC EMBL; U51388; AAB41693.1; -
 CC EMBL; U51389; AAB41694.1; -
 CC EMBL; U51390; AAB41695.1; -
 CC EMBL; U51391; AAB41696.1; -
 CC EMBL; U51392; AAB41697.1; -
 CC EMBL; U51393; AAB41698.1; -
 CC EMBL; U51394; AAB41699.1; -
 CC EMBL; U51395; AAB41700.1; -
 CC EMBL; U51396; AAB41701.1; -
 CC EMBL; U51397; AAB41702.1; -
 CC EMBL; U51398; AAB41703.1; -
 CC EMBL; U51399; AAB41704.1; -
 CC EMBL; U51400; AAB41705.1; -
 CC EMBL; U51401; AAB41706.1; -
 CC EMBL; U51402; AAB41707.1; -
 CC EMBL; U51403; AAB41708.1; -
 CC EMBL; U51404; AAB41709.1; -
 CC EMBL; U51405; AAB41710.1; -
 CC EMBL; U51406; AAB41711.1; -
 CC EMBL; U51407; AAB41712.1; -
 CC EMBL; U51408; AAB41713.1; -
 CC EMBL; U51409; AAB41714.1; -
 CC EMBL; U51410; AAB41715.1; -
 CC EMBL; U51411; AAB41716.1; -
 CC EMBL; U51412; AAB41717.1; -
 CC EMBL; U51413; AAB41718.1; -
 CC EMBL; U51414; AAB41719.1; -
 CC EMBL; U51415; AAB41720.1; -
 CC EMBL; U51416; AAB41721.1; -
 CC EMBL; U51417; AAB41722.1; -
 CC EMBL; U51418; AAB41723.1; -
 CC EMBL; U51419; AAB41724.1; -
 CC EMBL; U51420; AAB41725.1; -
 CC EMBL; U51421; AAB41726.1; -
 CC EMBL; U51422; AAB41727.1; -
 CC EMBL; U51423; AAB41728.1; -
 CC EMBL; U51424; AAB41729.1; -
 CC EMBL; U51425; AAB41730.1; -
 CC EMBL; U51426; AAB41731.1; -
 CC EMBL; U51427; AAB41732

FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO
 FT VARIANT 747 747 ALEGRE 1 AND PORTO ALEGRE 2).
 FT VARIANT 764 766 S -> A (IN STRAINS GUADELOUPE AND GUANA).
 FT VARIANT 886 886 MISSING (IN STRAIN MANAUS 3).
 FT NON_TER 1093 1093 A -> T (IN STRAIN 0811.4).
 SQ SEQUENCE 1093 AA; 115996 MW; AB6DE050267EC187 CRC64;
 Query Match 11.7%; Score 91; DB 1; Length 1093;
 Best Local Similarity 26.1%; Pred. No. 3.1;
 Matches 24; Conservative 8; Mismatches 14; Gaps 2;
 QY 27 GCGGNGGNGSGGPDYDQIVTRVTHMAHALQSDARKSETITQSGVNGADVGQGD 86
 DB 719 GCGGNGGNGSGGPGGGLPLFL-----DVTHTSSSQNGKPTGVAAGGAGGVGGGG-- 770
 QY 87 NSTIETONGFRNATIDQWNAKNKNSDITVGOY 118
 DB 771 -----SCSLGNGNGVSGNGNNSQPSINQY 796
 RESULT 11
 GP63 LEIMA STANDARD; PRT; 602 AA.
 AC P08148; P15906;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
 DE endopeptidase).
 GN GP63
 OS Leishmania major.
 OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Button L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of leishmania.";
 RL J. Exp. Med. 167:724-729(1988).
 RN [2]
 RP REVISIONS.
 RA Button L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589(1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
 RA Homans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 RT the Leishmania major promastigote surface protease.";
 RL J. Biol. Chem. 265:16955-16964(1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 RT leishmanolysin, the major surface metalloproteinase from Leishmania
 RT major.";
 RL Proteins 22:58-66(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 RT leishmanolysin.";
 RL Structure 6:1035-1046(1998).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.

CC -!- COPACTOR: Binds 1 zinc ion per subunit.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- PM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; Y00647; CAA68673.1; -.
 DR PIR; P0221; P0221.
 DR PDB; 1LML; 17-SEP-97.
 DR MEROPS; M08.001; -.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC PROTEASE; 1.
 DR KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 DR Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 100 ACTIVATION PEPTIDE.
 FT CHAIN 101 577 LEISHMANOLYSIN.
 FT PROPEP 578 602 REMOVED IN MATURE FORM.
 FT METAL 264 264 ZINC (CATALYTIC).
 FT ACT_SITE 265 265 ZINC (CATALYTIC).
 FT METAL 268 268 ZINC (CATALYTIC).
 FT METAL 334 334 ZINC (CATALYTIC).
 FT DISULFID 125 142
 FT DISULFID 191 230
 FT DISULFID 314 386
 FT DISULFID 393 455
 FT DISULFID 406 425
 FT DISULFID 415 489
 FT DISULFID 466 510
 FT DISULFID 515 565
 FT DISULFID 535 588
 FT CARBOHYD 300 300
 FT CARBOHYD 407 407
 FT LIPID 577 577
 FT STRAND 101 102
 FT STRAND 107 108
 FT STRAND 111 114
 FT HELIX 116 119
 FT TURN 121 122
 FT TURN 128 129
 FT STRAND 131 133
 FT STRAND 139 141
 FT HELIX 144 146
 FT HELIX 150 158
 FT TURN 159 159
 FT TURN 160 169
 FT TURN 170 171
 FT STRAND 172 174
 FT STRAND 177 178
 FT STRAND 180 181
 FT TURN 189 190
 FT HELIX 191 193
 FT HELIX 198 202
 FT TURN 203 203
 FT STRAND 205 206
 FT STRAND 210 215
 FT TURN 221 222
 FT STRAND 226 232
 FT TURN 234 235
 FT STRAND 238 244
 N-LINKED (GLCNAC. .) (POTENTIAL).
 N-LINKED (GLCNAC. .) (POTENTIAL).
 GPI-anchor amidated asparagine.

```

      247 249
      256 269
      270 271
      274 279
      280 281
      283 286
      289 291
      296 299
      302 312
      313 313
      315 316
      320 322
      328 332
      335 335
      337 339
      341 342
      343 343
      344 345
      353 353
      356 364
      365 366
      369 370
      372 374
      380 383
      386 390
      395 395
      396 397
      398 399
      402 404
      413 414
      417 418
      421 425
      428 429
      435 437
      443 444
      445 446
      450 454
      458 465
      466 467
      470 472
      475 477
      480 480
      485 486
      487 494
      496 496
      506 516
      517 520
      521 525
      527 528
      533 534
      537 538
      540 542
      543 545
      546 546
      550 550
      552 553
      555 557
      561 565
      566 567
      569 572
      573 573
SQ  SEQUENCE 602 AA; 982EF3245D87C43E CRC64;

      11.5%; Score 90; DB 1; Length 602;
      Best Local Similarity 33.3%; Pred. No. 1.9;
      Matches 28; Conservative 6; Mismatches 16; Indels 34; Gaps 3;

      43 YDOLVTRVTHVTHAHLQ-----SDAR-----KSETTITQSGYNG 78
      254 YDOLVTRVTHVTHAHLGSGFFEDARIVANVPVRGKNFVVPVINSSTAVAKREQYG 313
      79 ADV-----GQADNSTIEL 92
      |::|::|

```

```

DR MM, 138248; -.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0008066; F:glutamate receptor activity; TAS.
DR GO; GO:0015277; F:kainate selective glutamate receptor activity; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR GO; GO:0007268; P:synaptic transmission; TAS.
DR InterPro; IPRO01828; ANF receptor.
DR InterPro; IPRO01320; Ion glu receptor.
DR InterPro; IPRO01622; K+channel pore.
DR InterPro; IPRO01508; NMDA receptor.
DR InterPro; IPRO01311; SBP/Glu receptor.
DR Pfam; PF01094; ANF_receptor; 1.
DR Pfam; PF00060; lig_chan; 1.
DR PRINTS; PR00177; NMDARECEPTOR.
DR SMART; SM00079; PBP; 1.
KW Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
Transmembrane; Alternative splicing.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 906 GLUTAMATE RECEPTOR 1.
FT DOMAIN 19 536 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 537 557 POTENTIAL.
FT TRANSMEM 585 605 POTENTIAL.
FT TRANSMEM 618 638 POTENTIAL.
FT TRANSMEM 806 826 POTENTIAL.
FT VARSPIC 758 758 N -> G (in isoform Flip).
FT VARSPIC 768 768 N -> S (in isoform Flip).
FT VARSPIC 772 772 L -> V (in isoform Flip).
FT VARSPIC 778 778 N -> S (in isoform Flip).
FT VARSPIC 790 793 GSGD -> KDSG (in isoform Flip).
FT CONFLICT 345 345 R -> A (in REF. 2).
FT CONFLICT 375 375 S -> G (in REF. 2 AND 3).
FT CONFLICT 863 863 S -> SA (in REF. 3).
FT CONFLICT 865 867 AGA -> TAP (in REF. 3).
FT SEQUENCE 906 AA; 101536 MW; 03EAL06D1D0CE8EF CRC64;
Query Match 11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 3;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;
QY 6 VAAFAIVVSGALGVVPWQGGGNGHGGNSGPDYDOL-----VTRVVTHE---M 55
Db 250 VTGFQLVNVTDTIPAKINQW-----KNSDARDHTRVDWKPXYTSALTYDGVKVM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----CQGAD-----NSTIELTQ 94
Db 301 AEAFQSLRQ---RIDISRRGNAGDCLANPAVFWQGGIDICRALQOVAFEGLTGNVQNE 357
QY 95 NGRFNAT-----IDQNAKNSDI---TVGYGNNALVNQ-----ASD 132
Db 358 KERRNYTLHVTEMKHSIRKGYWNEDDKFPAATDAQAGDDNSQVNRITVITILED 417
QY 133 SSMVVRQVFGNNATANQY 151
Db 418 PYVWLKK-----NANQF 429
RESULT 13
ID OMPB_RICCN STANDARD; PRT; 1655 AA.
AC Q9KXK3; Q9KXK3; Q9XC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (surface protein antigen) (Cell surface antigen 5) (Scas) (rOMPb)
DE (rOMP B) [contains: 120 kDa surface-exposed protein (surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB OR RC1085.

```

```

OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiidae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RT Science 293:2093-2098(2001).
[2]
SEQUENCE OF 33-1649 FROM N.A.
RC STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V., Raoult D.;
RA "Phylogenetic analysis of members of the genus Rickettsia using the
RA gene coding the outer-membrane protein ompB (ompB).";
RT Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
[3]
SEQUENCE OF 353-1655 FROM N.A.
RC STRAIN=Malish 7;
RA Stenos J., Walker D.;
RA "The rickettsial outer membrane protein A and B genes of Rickettsia
RA australis, the most divergent rickettsia of the spotted fever group.";
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008659; AAL03623.1; -.
CC EMBL; AF123721; AAF34124.1; -.
CC EMBL; AF123726; AAF34129.1; -.
CC EMBL; AF149110; AAD39533.1; -.
CC PIR; E97835; E97835.
CC InterPro; IPRO06315; Autotransport.
CC Pfam; PF03797; Autotransporter.
CC TIGRFAMS; TIGR01414; autotrans_1.
CC Antigen; S-layer; Cell wall; Complete proteome.
KW CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1335 1655 32 kDa BETA PEPTIDE.
FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT CONFLICT 353 354 KD -> GH (IN REF. 3).
FT CONFLICT 776 776 P -> S (IN REF. 3).
FT CONFLICT 1159 1159 E -> D (IN REF. 3).
FT CONFLICT 1177 1177 G -> S (IN REF. 3).
FT CONFLICT 1492 1492 H -> R (IN REF. 3).
FT SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

```

| | |
|----------------|--|
| DR | PROSITE; PS00142; ZINC_PROTEASE; 1. |
| KW | Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc; |
| KW | Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein. |
| FT | SIGNAL 1 39 POTENTIAL. |
| FT | PROPEP 40 97 ACTIVATION PEPTIDE. |
| FT | CHAIN 98 574 LEISHMANOLYSIN. |
| FT | PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY). |
| FT | METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY). |
| FT | ACT_SITE 262 262 BY SIMILARITY. |
| FT | METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY). |
| FT | METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY). |
| FT | DISULFID 122 139 BY SIMILARITY. |
| FT | DISULFID 188 227 BY SIMILARITY. |
| FT | DISULFID 311 383 BY SIMILARITY. |
| FT | DISULFID 390 452 BY SIMILARITY. |
| FT | DISULFID 403 422 BY SIMILARITY. |
| FT | DISULFID 412 486 BY SIMILARITY. |
| FT | DISULFID 463 507 BY SIMILARITY. |
| FT | DISULFID 512 562 BY SIMILARITY. |
| FT | CARBHYD 532 555 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT | CARBHYD 394 394 GPI-anchor amidated asparagine (By |
| FT | LIPID 574 574 similarity) |
| SQ | SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64; |
| | Query Match 11.3%; Score 88; DB 1; Length 599; |
| | Best Local Similarity 46.7%; Pred. No. 2.8; |
| | Matches 21; Conservative 3; Mismatches 11; Indels 10; Gaps 1; |
| QY | 43 YDQLVTRVVTHMAHALQSARKSETITQSGYGNGADVCGGADN 87 |
| DB | 251 YDQLVTRVVTHMAHALGFVS-----GFEGARILEISGN 285 |
| RESULT 15 | |
| OVO_DROME | STANDARD; PRT; 1028 AA. |
| ID - OVO DROME | |
| AC | P51521; Q9XZU4; |
| DT | 01-OCT-1996 (Rel. 34, Created) |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) |
| DE | Ovo protein (Shaven baby protein). |
| OS | OVO OR SVE. |
| GN | Drosophila melanogaster (Fruit fly). |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; |
| OC | Nectera; Endopterygota; Diptera; Brachycera; Muscomorpha; |
| OC | Ochdroidea; Drosophilidae; Drosophila. |
| OX | NCBI_TaxID=7227; |
| RN | [1] |
| RC | SEQUENCE FROM N.A. |
| RP | TISSUE=Ovary; |
| RA | MEDLINE=35031209; PubMed=7935398; |
| RX | Gadginkal M.D., Wang J., Liang Y., Mahowald A.P.; |
| RT | "Multiple products from the shavenbaby-ovo gene region of Drosophila |
| RT | melanogaster: relationship to genetic complexity."; |
| RL | Mol. Cell. Biol. 14:6809-6818(1994). |
| RN | [2] |
| RC | SEQUENCE FROM N.A. |
| RP | STRAIN=Oregon-R; |
| RX | MEDLINE=91293102; PubMed=1712294; |
| RT | Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.; |
| RT | "The ovo gene of Drosophila encodes a zinc finger protein required |
| RT | for female germ line development."; |
| RL | EMBO J. 10:2259-2266(1991). |
| CC | -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM |
| CC | LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION. |
| CC | -!- SUBCELLULAR LOCATION: Nuclear (Potential). |
| CC | -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARINUM AND |
| CC | ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG, |
| CC | BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED |
| CC | PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS. |
| CC | -!- SIMILARITY: Contains 4 C2H2-type zinc fingers. |

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; U11383; AAB60216.1; -
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSSP; P07248; 2ADR.
DR TRANSFAC; T00669; -
DR FlyBase; FBgn0003028; ovo.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 823 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT CONFLICT 647 647 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;

Query Match 11.2%; Score 87.5; DB 1; Length 1028;
Best Local Similarity 25.6%; Pred No. 5.7;
Matches 40; Conservative 11; Mismatches 58; Indels 47; Gaps 6;
QY 3 LLKVAFAAIVVSGSALAGVFWGCGGNHNGGNSGPDYDLVTRVVTHEMAHALQSD 62
Db 59 LQNAAAAAYIMGASG-----GGGCTGGGGGASGP----- 89
QY 63 ARKSETTITQSGYGNAGVGGADN-----STIELTONGFNNATIDQWAKNSDI--- 113
Db 90 ----GGGPSANSGGGGGGGGNGYINCGGVGFPNNSLDGNLLNFASVSNYESNSKFNH 146
QY 114 -TVGYGGNNAALVNQTASDSSVMVQVGFGNATA 148
Db 147 HHHQHNNNN-----NNGGQTSMMGHFP--YGNPSPA 177

Search completed: August 2, 2004, 14:49:29
Job time : 5.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-18
Perfect score: 780
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 682 | 87.4 | 152 | 2 Q33802 | O33802 salmonella |
| 2 | 593.5 | 76.1 | 150 | 2 Q7X243 | Q7X243 citrobacter |
| 3 | 550 | 70.5 | 149 | 2 Q7X240 | Q7X240 citrobacter |
| 4 | 498.5 | 63.9 | 152 | 16 Q8CW63 | Q8CW63 escherichia |
| 5 | 428.5 | 54.9 | 150 | 2 Q7X237 | Q7X237 enterobacte |
| 6 | 303 | 38.8 | 76 | 2 Q54069 | Q54069 salmonella |
| 7 | 122 | 15.6 | 29 | 2 Q9S3J5 | Q9S3J5 escherichia |
| 8 | 114 | 14.6 | 1748 | 5 Q94821 | Q94821 tetrahymena |
| 9 | 113 | 14.5 | 502 | 16 Q8EIH4 | Q8EIH4 shewanella |
| 10 | 110.5 | 14.2 | 151 | 2 Q7X244 | Q7X244 citrobacter |
| 11 | 108 | 13.8 | 171 | 16 Q89J13 | Q89J13 bradyrhizob |
| 12 | 107.5 | 13.8 | 151 | 16 Q7UC21 | Q7UC21 shigella fl |
| 13 | 107.5 | 13.8 | 160 | 16 Q8CW64 | Q8CW64 escherichia |
| 14 | 107.5 | 13.8 | 160 | 16 Q83RU7 | Q83RU7 shigella fl |
| 15 | 106.5 | 13.7 | 1209 | 16 Q89CK5 | Q89CK5 bradyrhizob |
| 16 | 105.5 | 13.5 | 91 | 2 Q9S3J8 | Q9S3J8 escherichia |

| | | | | | |
|----|-------|------|------|-----------|--------------------|
| 17 | 101.5 | 13.0 | 154 | 16 Q89J15 | Q89J15 bradyrhizob |
| 18 | 101 | 12.9 | 262 | 5 Q9VIX5 | Q9VIX5 drosophila |
| 19 | 100 | 12.8 | 179 | 2 Q33801 | O33801 salmonella |
| 20 | 99 | 12.7 | 130 | 16 Q89J14 | Q89J14 bradyrhizob |
| 21 | 98 | 12.6 | 157 | 16 Q88HG0 | Q88HG0 pseudomonas |
| 22 | 98 | 12.6 | 362 | 16 Q8EV84 | Q8EV84 mycoplasma |
| 23 | 98 | 12.6 | 3659 | 16 Q98LN6 | Q98LN6 rhizobium 1 |
| 24 | 97.5 | 12.5 | 151 | 2 Q7X238 | Q7X238 enterobacte |
| 25 | 97.5 | 12.5 | 152 | 2 Q7X241 | Q7X241 citrobacter |
| 26 | 97 | 12.4 | 368 | 16 Q8EWD6 | Q8EWD6 mycoplasma |
| 27 | 97 | 12.4 | 1422 | 16 Q8EFU3 | Q8EFU3 shewanella |
| 28 | 97 | 12.4 | 2035 | 2 Q9XCJ4 | Q9XCJ4 salmonella |
| 29 | 97 | 12.4 | 2039 | 16 Q8ZN57 | Q8ZN57 salmonella |
| 30 | 96.5 | 12.4 | 1765 | 16 Q7V855 | Q7V855 prochloroco |
| 31 | 96 | 12.3 | 490 | 16 Q8EY19 | Q8EY19 leptospira |
| 32 | 95.5 | 12.2 | 145 | 16 Q8U6N9 | Q8U6N9 agrobacteri |
| 33 | 95.5 | 12.2 | 1613 | 2 Q9KKB2 | Q9KKB2 israeli tic |
| 34 | 95.5 | 12.2 | 3501 | 16 Q8Y106 | Q8Y106 ralstonia s |
| 35 | 95.5 | 12.2 | 3552 | 16 Q8XSD6 | Q8XSD6 ralstonia s |
| 36 | 94.5 | 12.1 | 153 | 16 Q89J16 | Q89J16 bradyrhizob |
| 37 | 94.5 | 12.1 | 348 | 13 Q93397 | Q93397 cyprinus ca |
| 38 | 94 | 12.1 | 480 | 16 Q89EV2 | Q89EV2 bradyrhizob |
| 39 | 93.5 | 12.0 | 139 | 16 Q8EIH3 | Q8EIH3 shewanella |
| 40 | 93.5 | 12.0 | 287 | 5 Q9VIX6 | Q9VIX6 drosophila |
| 41 | 93.5 | 12.0 | 582 | 16 P71868 | P71868 mycobacteri |
| 42 | 93.5 | 12.0 | 582 | 16 Q7TW98 | Q7TW98 mycobacteri |
| 43 | 93.5 | 12.0 | 1615 | 2 Q9KKA8 | Q9KKA8 rickettsia |
| 44 | 93 | 11.9 | 141 | 16 Q8U6P1 | Q8U6P1 agrobacteri |
| 45 | 92 | 11.8 | 353 | 16 Q8EV92 | Q8EV92 mycoplasma |

ALIGNMENTS

RESULT 1

O33802 ID O33802 PRELIMINARY; PRT; 152 AA.
AC O33802;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Agfa protein (Fragment).
GN AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells".
RL Infect. Immun. 65:5320-5325 (1997).
DR EMBL; AJ000514; CAA04151.1; -.
FT NON_TER 152 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.4%; Score 682; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 1.7e-46;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDYDQLVTRVVTWTHMAHALQ 60
1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYGYGSANAALAQ 60
DB 61 SPARKSETTITSGYNGADVCGGADNSTIETQTGFRNATIDOWNAKNSDITVQYCG 120
61 SPARKSETTITSGYNGADVCGGADNSTIETQTGFRNATIDOWNAKNSDITVQYCG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANYQ 151

```

Db      121 NNAALVNQTASDSSVMVRQVGFNNAPANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2.
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 76.1%; Score 593.5; DB 2; Length 150;
Best Local Similarity 78.1%; Pred. No. 1.7e-39;
Matches 118; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60
|||||
Db 1 MKLLQVAAPAAIIVSGSALAGVVPQWGGGG-GGGSSSGPESTLSIYQSGVNAALAQ 59
|||||
QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
|||||
Db 60 SPARKSDTIHQNGFGNGADVCGGSDNSTIDLQNGFKNNATIDQWNGKNSDITVQYGG 119
|||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 70.5%; Score 550; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 4.5e-36;
Matches 109; Conservative 20; Mismatches 20; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60
|||||

```

```

Db      1 MKLLKVAAPAAIIVSGSALAGVVPQW--GGNHGGGNSYGPDSLSIYQYGSNNSANALQ 58
QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
|||||
Db 59 SPARKSDVTITQHRGNGAVVQGGADDSISLQGTGFQNSATIDQWNAKNSDITVQFGG 118
|||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 119 RNGALVNQTASDSNVLIQQVGFNNATANQH 149
|||||

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AC016759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 63.9%; Score 498.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 5.4e-32;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQW-GGGNGHGGGNSGGPDYDQLVTRVVTHEMAHAL 59
|||||
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQYGGGNGHGGGNSGNSBELNIYQYGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 119
|||||
Db 61 QADARNSDLTITQHGNGADVCGGSDSSIDLQNGFGNSATLDQWNGKNSDITVQKQFG 120
|||||
QY 120 GNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 121 GNGAARDQTSNSSVNVTVQVGFNNATAHQY 152
|||||

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

```



```
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8B32D872DF1593 CRC64;

Query Match 54.9%; Score 428.5; DB 2; Length 150;
Best Local Similarity 55.1%; Pred. No. 1.8e-26;
Matches 92; Conservative 21; Mismatches 21; Indels 33; Gaps 3;

QY 1 MLLKVAFAAIVVSGSALAGVVPQ--WGGGNNH-----NGGKSSGPDYD 44
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MKFKVAALAAIVVSGSMAGINQGGWGHGHGGVGGPNTLIYQNGGNS----- 54

QY 45 QLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVCGQAGNDSITELTQNGFRNNATID 104
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 55 -----ALALQTDARNSVLNISQTGGNGADVCGQSDSSINLTQNGFGNSATLD 103

QY 105 QWNAKNSDITVQYGGNNALAVNOTASDSSVMVQVQVGFNNATANOY 151
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 104 QWNSKDSVMVNSQYGGNGALVQDQASNSTVNVTTQIGFNGHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RA "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBSJ databases.
DR EMBL: U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.8%; Score 303; DB 2; Length 76;
Best Local Similarity 80.3%; Pred. No. 6.6e-17;
Matches 61; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 30 GNHNGGSSGPDYDQVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGGADNST 89
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 GNHNGGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNST 60

QY 90 IELTQNGFRNNATIDQ 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q983J5 PRELIMINARY; PRT; 29 AA.
ID Q983J5;
AC Q983J5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-WAR-2001 (TREMBlrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
```

```
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOS-Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Regione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:K80 isolates associated with
RT ISI inserti on in cs9B and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.6%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0042;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGG 29
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MLLKVAALAAIVVSGSALAGVVPQWGGG 29

RESULT 8
Q94821 PRELIMINARY; PRT; 1748 AA.
ID Q94821;
AC Q94821; P92146; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CNJB protein.
GN CNJB.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=3357771;
RA Martindale D.W., Taylor F.M.;
RA "Multiple introns in a conjugation-specific gene from Tetrahymena
RT thermophila.";
RL Nucleic Acids Res. 16:2189-2201(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94051569; PubMed=8233798;
RA Taylor F.M., Martindale D.W.;
RA "Retroviral-type zinc fingers and glycine-rich repeats in a protein
RT encoded by cnjB, a Tetrahymena gene active during meiosis.";
RL Nucleic Acids Res. 21:4610-4614(1993).
DR EMBL: X06462; CAB37323.1; -.
DR EMBL: L03710; AAC37171.1; -.
DR PIR: S42136; S42136.
DR HSSP: P85888; IAAF.
DR GO: G0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf_CCHC; 7.
DR PRINTS: PR00939; C2HGCZNFINGER.
DR SMART: SM00343; Znf_C2HC; 7.
DR PROSITE: PS00158; ZF_CCHC; 7.
FT CONFLICT 251 M -> I (IN REF. 1).
FT CONFLICT 256 I -> N (IN REF. 1).
FT CONFLICT 256 I -> N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 14.6%; Score 114; DB 5; Length 1748;
Best Local Similarity 30.5%; Pred. No. 2.3;
Matches 39; Conservative 22; Mismatches 27; Indels 40; Gaps 8;

QY 25 QWGGGNNHNGG---GNSSGPDYDQVTRVVTHEMAHALQSDARKSETTITQSGYNGADV 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1640 QFGGGNSNGSGSWGTSSGSDWN-----CGSNVQES-TTSSGGWGS----- 1680

QY 82 GQAGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNALAVNOTASDSSVMVQV 141
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

Db 1681 -SGSGNQ-----TGGWGSN---DNQQQQQVNTGGGWSNS---NQTNSS----- 1722
Qy 142 FGNNTAN 149
Db 1723 WGSNNQAS 1730

RESULT 9
Q8EIH4 PRELIMINARY; PRT; 502 AA.
ID Q8EIH4
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SC0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadales; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2297686; PubMed=12368913;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsipin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Nelson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AF015532; RAN53941.1; -.
DR TIGR; S00865; -.
SQ HYPOTHETICAL protein; Complete proteome.

Query Match 14.5%; Score 113; DB 16; Length 502;
Best Local Similarity 26.0%; Pred. No. 0.64;
Matches 40; Conservative 20; Mismatches 58; Indels 36; Gaps 6;

Qy 29 GGNHNG-----GGN-----SSGPDYDQLVTR-----VYTHEMAHAL 59
Db 231 GDNHTGVYALAGSNDISMEOEGSNNTAYLSMTTGDNTVDITDGDSTNTVGSLSIADI 290

Qy 60 QSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQGYG 119
Db 291 QGD--DNDITIKQKDSNGAEQFVWGSNDVDLKGQDANFATFGAYGTDN-DFDLSKSG 347

Qy 120 GNNALVNQTASDSSVMVRQVGFN-----NATAN 149
Db 348 DNNELVAFATGDNSEIISQEGDANFAYVDATGN 381

RESULT 10
Q7X244 PRELIMINARY; PRT; 151 AA.
ID Q7X244
AC Q7X244;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;

```

```

RA Zogaj X., Bokranz W., Nintz M., Ronling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AU515700; CAD56671.1; -.
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.2%; Score 110.5; DB 2; Length 151;
Best Local Similarity 25.0%; Pred. No. 0.24;
Matches 29; Conservative 26; Mismatches 46; Indels 15; Gaps 3;

Qy 47 VTRVVTHEMAHALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFR----- 98
Db 18 IASATSYDLAH---SEYNFAVNELSKSFNQAIIGQVGTNNSAKVRQSGKLLSVVSGE 74
Qy 99 ---NNATIDOWNAKNSDITVQGVGGNNAALVNQTASDSSVMVRQVGFNNTANQY 151
Db 75 GGSNRAKVDSGAYNF-AVIAQSGHNSDASISQSNVYGTAMIIQKSGNKANITQY 129

RESULT 11
Q89JI3 PRELIMINARY; PRT; 171 AA.
ID Q89JI3
AC Q89JI3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CSGA protein.
GN CSGA OR BLL5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 13.8%; Score 108; DB 16; Length 171;
Best Local Similarity 36.6%; Pred. No. 0.44;
Matches 34; Conservative 10; Mismatches 39; Indels 10; Gaps 3;

Qy 58 ALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITV 115
Db 52 AUSSAAQAATSTITVQVGLVNGSSVTQNGLTNDSSSTTQIGILNGASTVQGTSSFS---- 107
Qy 116 GOYGGNNAALVNQTASDSSVMVRQVGFNNTA 148
Db 108 ----LNNVSTVQAGVQNSATTGQVAFGNGSA 136

RESULT 12
Q7UCZ1 PRELIMINARY; PRT; 151 AA.
ID Q7UCZ1
AC Q7UCZ1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.

```

```

ON NCBI_TaxID=623;
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=2259274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016981; AAP16542.1; -.
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.42;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
Db 49 IGQAGTNSAQLRGGSKLLAVVAQEGSSNRKIDQTGDYNL-AYIDQAGSANDASISQ 107

QY 130 ASDSSVMVRQVGFNNATANY 151
Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 13
Q8CW64
ID Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1];
RP SEQUENCE FROM N.A.
RX STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
Db 58 IGQAGTNSAQLRGGSKLLAVVAQEGSSNRKIDQTGDYNL-AYIDQAGSANDASISQ 116

QY 130 ASDSSVMVRQVGFNNATANY 151
Db 117 AYGNTAMIIQKSGNKANITQY 138

RESULT 14
Q83RU7
ID Q83RU7 PRELIMINARY; PRT; 160 AA.
AC Q83RU7;
```

```

DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to CSga.
GN CSGB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1];
RP SEQUENCE FROM N.A.
RX STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AB015131; AAN42658.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45;
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQT 129
Db 58 IGQAGTNSAQLRGGSKLLAVVAQEGSSNRKIDQTGDYNL-AYIDQAGSANDASISQ 116

QY 130 ASDSSVMVRQVGFNNATANY 151
Db 117 AYGNTAMIIQKSGNKANITQY 138

RESULT 15
Q89CK5
ID Q89CK5 PRELIMINARY; PRT; 1209 AA.
AC Q89CK5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BIL7792 protein.
GN BIL7792;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1];
RP SEQUENCE FROM N.A.
RX STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasanoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005963; BAC53057.1; -.
KW Complete proteome.
SQ SEQUENCE 1209 AA; 118292 MW; 50F97581D524EB71 CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 1209;
Best Local Similarity 25.0%; Pred. No. 5.9;
Matches 47; Conservative 20; Mismatches 66; Indels 55; Gaps 8;

QY 9 FRAIVVGSALAGVVQWGGGNGHNGGSSCPDYLQLVTRVVTHEMAH----ALQSDAR 64
Db 934 YGNVTVGTGTGATTIAL----GNGNDTIDASG--YGNVITLNGNDIVHPGDGASQTAG 987
```

| | | | |
|----|------|---|------|
| QY | 65 | KSETTITQSGYNGADVGQAD-----NSTIELTQNGFRNNATIDQWN- | 107 |
| Db | 988 | NGNDLVTLGSGYNTVMLGNGNDVWAGRDGANSVTILGDGNTVNL--GGMGNIITVGSQT | 1045 |
| QY | 108 | -----AKNSDITVQYQ-----GNNALVNO-----TASDSSVMVRQVG | 141 |
| Db | 1046 | ALIAGSGSDNVVAGAGHDTIMLGGAAHNVVLNGSQAVTNIQGDVVTVNGGSDQFNFG | 1105 |
| QY | 142 | FCNNATAN | 149 |
| Db | 1106 | FCNQAIIN | 1113 |

Search completed: August 2, 2004, 14:54:38
Job time : 29.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVFGNNATNQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 774 | 100.0 | 151 | 3 AAB36350 | Aab36350 Agfa::PT3 |
| 2 | 709 | 91.6 | 151 | 3 AAB36353 | Aab36353 Agfa::PT3 |
| 3 | 689 | 89.0 | 151 | 2 AAR74625 | Aar74625 Agfa sequ |
| 4 | 689 | 89.0 | 151 | 3 AAB36341 | Aab36341 Salmonell |
| 5 | 684 | 88.4 | 151 | 2 AAW23570 | Aaw23570 Salmonell |
| 6 | 662 | 85.5 | 151 | 3 AAB36354 | Aab36354 Agfa::PT3 |
| 7 | 622 | 80.4 | 151 | 3 AAB36349 | Aab36349 Agfa::PT3 |
| 8 | 619 | 80.0 | 151 | 3 AAB36351 | Aab36351 Agfa::PT3 |
| 9 | 611 | 78.9 | 151 | 3 AAB36346 | Aab36346 Agfa::PT3 |
| 10 | 609 | 78.7 | 151 | 3 AAB36347 | Aab36347 Agfa::PT3 |
| 11 | 606 | 78.3 | 151 | 3 AAB36352 | Aab36352 Agfa::PT3 |
| 12 | 597 | 77.1 | 151 | 3 AAB36355 | Aab36355 Agfa::PT3 |
| 13 | 574 | 74.2 | 151 | 3 AAB36348 | Aab36348 Agfa::PT3 |
| 14 | 528 | 68.2 | 151 | 3 AAB36343 | Aab36343 Escherich |
| 15 | 523 | 67.6 | 151 | 7 ABR82651 | Abr82651 E. coli C |
| 16 | 504 | 65.1 | 120 | 2 AAR62761 | Aar62761 Agfa sequ |
| 17 | 504 | 65.1 | 120 | 2 AAW23569 | Aaw23569 Salmonell |
| 18 | 450 | 58.1 | 142 | 2 AAR52664 | Aar52664 Fibronect |
| 19 | 378 | 48.8 | 122 | 2 AAR52663 | Aar52663 FNB curli |
| 20 | 153 | 19.8 | 45 | 3 AAB36316 | Aab36316 Salmonell |
| 21 | 132 | 17.1 | 22 | 3 AAB36318 | Aab36318 Salmonell |
| 22 | 123 | 15.9 | 23 | 3 AAB36321 | Aab36321 Salmonell |
| 23 | 123 | 15.9 | 23 | 3 AAB36326 | Aab36326 Salmonell |
| 24 | 123 | 15.9 | 23 | 3 AAB36338 | Aab36338 Salmonell |
| 25 | 113 | 14.6 | 24 | 7 ABR82644 | Abr82644 E. coli C |

| | | | | | |
|----|------|------|-----|------------|--------------------|
| 26 | 111 | 14.3 | 22 | 3 AAB36322 | Aab36322 Salmonell |
| 27 | 111 | 14.3 | 22 | 3 AAB36327 | Aab36327 Salmonell |
| 28 | 111 | 14.3 | 22 | 3 AAB36337 | Aab36337 Salmonell |
| 29 | 109 | 14.1 | 23 | 3 AAB36340 | Aab36340 Salmonell |
| 30 | 109 | 14.1 | 23 | 3 AAB36324 | Aab36324 Salmonell |
| 31 | 109 | 14.1 | 23 | 3 AAB36319 | Aab36319 Salmonell |
| 32 | 102 | 13.2 | 26 | 7 ABR82649 | Abr82649 E. coli V |
| 33 | 96.5 | 12.5 | 151 | 3 AAB36344 | Aab36344 Escherich |
| 34 | 96 | 12.4 | 19 | 3 AAB36323 | Aab36323 Salmonell |
| 35 | 96 | 12.4 | 19 | 3 AAB36336 | Aab36336 Salmonell |
| 36 | 96 | 12.4 | 19 | 3 AAB36328 | Aab36328 Salmonell |
| 37 | 95 | 12.3 | 24 | 7 ABR82647 | Abr82647 E. coli c |
| 38 | 92 | 11.9 | 23 | 3 AAB36331 | Aab36331 Escherich |
| 39 | 91 | 11.8 | 186 | 6 ABU21488 | Abu21488 Protein e |
| 40 | 91 | 11.7 | 502 | 2 AAW32312 | Aaw32312 Leishmani |
| 41 | 90.5 | 11.7 | 677 | 4 AAG04318 | Ag04318 Novel hum |
| 42 | 90 | 11.6 | 24 | 7 ABR82642 | Abr82642 E. coli N |
| 43 | 89.5 | 11.6 | 423 | 4 AAG07164 | Ag07164 Novel hum |
| 44 | 89.5 | 11.6 | 447 | 3 AAG29728 | Aag29728 Arabidops |
| 45 | 89.5 | 11.6 | 468 | 3 AAG29727 | Aag29727 Arabidops |

ALIGNMENTS

RESULT 1

AAB36350

ID AAB36350 standard; protein; 151 AA.

XX

AC AAB36350;

XX

DT 26-FEB-2001 (first entry)

XX

DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.

XX

KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

XX

KW vaccine; immune response; immunogen.

XX

OS Salmonella enteritidis.

OS

OS Escherichia coli.

OS

XX Synthetic.

XX

PN WO2000060102-A2.

XX

PD 12-OCT-2000.

XX

PF 05-APR-2000; 2000WO-CA000356.

XX

PR 05-APR-1999; 99US-0127888P.

XX

XX (UYVI-) UNIV VICTORIA.

PI

White AP, Doxan JL, Collison SK, Kay WW;

XX

WPI: 2000-672631/65.

DR

N-PSDB; AAC64826.

XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

XX

protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 137; 139pp; English.

CC

The present invention describes a recombinant agfa gene (I) where a

CC

segment of the gene has been replaced by a segment of a foreign DNA

CC

sequence which encodes a foreign epitope or antigen. Also described are:

CC

(1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended

CC

assembly system of strains of Salmonella, Escherichia coli and

CC

Enterobacteriaceae for the production of fimbriae comprising recombinant

CC

Agfa, CSga and Agfa-homologue fimbrial subunits, respectively; (2)

CC

directing recombination of a recombinant gene into the chromosome of the

CC

homologous species; (3) directing recombination of a recombinant gene

CC

back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 100.0%; Score 774; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 8.8e-71;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 QY 61 SDARKYDQLVTRVVTHEMAHAGGAGDNSTIELTQGNFRNATIDQWNAKNSDIIVGOYGG 120
 DB 61 SDARKYDQLVTRVVTHEMAHAGGAGDNSTIELTQGNFRNATIDQWNAKNSDIIVGOYGG 120
 QY 121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151

RESULT 2
 AAB36353
 ID AAB36353 standard; protein; 151 AA.
 AC AAB36353;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 DE
 DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 XX WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64629.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 139; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, Cga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 91.6%; Score 709; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 3.7e-64;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALQ 57
 QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIELTQGNFRNATIDQWNAKNSD 112
 DB 58 -----YDQLVTRVVTHEMAHAGYNGADVGQADNSTIELTQGNFRNATIDQWNAKNSD 112
 QY 113 ITVGQYGGNNAALVNOTASDSSVWVRQVGFNNATANQY 151
 DB 113 ITVGQYGGNNAALVNOTASDSSVWVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 XX AAR74625;
 AC
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX Agfa sequence.
 DE
 XX Salmonella; Agfa; vaccine.
 KW
 XX Salmonella.
 OS
 XX WO9425598-A2.
 PN
 XX 10-NOV-1994.
 PD
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compans. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The Salmonella AgfA protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 151 AA;
 Query Match 89.0%; Score 689; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIETQGFNRNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIETQGFNRNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 4
 AAB36341 AAB36341 standard; protein; 151 AA.
 XX
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrian; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli, and
 CC enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CSgA and AgfA-homologue fimbrian subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrian protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 89.0%; Score 689; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIETQGFNRNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIETQGFNRNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 5
 AAW23570 AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfA.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX Collinson SK, Kay WW, Doran JL;
 PI

```

XX WPI; 1997-309886/28.
XX DR N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfa encoded by the full agfa gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteriaceae. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridise to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
SQ
Query Match 88.4%; Score 684; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.3e-61;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIIWVSGSALAGVVPQGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAAIIWVSGSALAGVVPQGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQGDNSTIETQNGFRNNATIDQWNAKNSDIITVGQYGG 120
DB 61 SDARKSETTITQSGYGNCGADYDQLVTRVVTHEMAHAGQGDNSTIETQNGFRNNATIDQWNAKNSDIITVGQYGG 120
QY 121 NNAALVNTQASDSSVMVQVGFNNATANOY 151
DB 121 NNPALVNTQASDSSVMVQVGFNNATANOY 151
RESULT 6
AAB36354
ID AAB36354 standard; protein; 151 AA.
AC AAB36354;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/55.
XX DR N-PSDB; AAC64630.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
Query Match 85.5%; Score 662; DB 3; Length 151;
Best Local Similarity 81.3%; Pred. No. 2.3e-59;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLLKVAFAAIIWVSGSALAGVVPQGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAAIIWVSGSALAGVVPQGGGNGHNGGSGDPSTLSIYQGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQGDNSTIETQNGFRNNATIDQ 105
DB 61 SDARKSETTITQSGYGNCGADYDQLVTRVVTHEMAHAGQGDNSTIETQNGFRNNATIDQ 105
QY 106 WNAKNSDIITVGQYGNNAALVNTQASDSSVMVQVGFNNATANOY 151
DB 106 WNAKNSDIITVGQYGNNAALVNTQASDSSVMVQVGFNNATANOY 151
RESULT 7
AAB36349
ID AAB36349 standard; protein; 151 AA.
AC AAB36349;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX
XX

```


AAB36346
ID AAB36346 standard; protein; 151 AA.
XX AAB36346;
AC AAB36346;
XX AAB36346;
DT 26-FEB-2001 (first entry)
XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
DE AAB36346 standard; protein; 151 AA.
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX Disclosure; Page 135; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (Srf17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 78.9%; Score 611; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 3.7e-54;
XX Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
XX
XX 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
XX 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 78.7%; Score 609; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.8e-54;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120
QY 121 NNAALVNOTASDVSVVRQVGFNNATANQY 151
DB 121 LVTRVTHEMAHASVVMVRQVGFNNATANQY 151

RESULT 11
AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
DT 26-FEB-2001 (first entry)
DE AgfA::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64628.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine; the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 78.3%; Score 606; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.2e-53;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVTRVVT 120
QY 121 NNAALVNOTASDVSVVRQVGFNNATANQY 151
DB 121 HEMAHANQTASDVSVVRQVGFNNATANQY 151

RESULT 12
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
DT 26-FEB-2001 (first entry)
DE AgfA::PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64631.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.
XX
CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 77.1%; Score 597; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 9.8e-53;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWVGNGGNGSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWVGNGGNGSGDPSTLSIYQGSANAALALQ 60

QY 61 SPARKYDQLVTRVVTHEMAHAGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120
 DB 61 SPARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAHAGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX
 AC AAB36348;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.
 XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;
 SQ

Query Match 74.2%; Score 574; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.2e-50;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPWVGNGGNGSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSGDPSTLSIYQGSANAALALQ 60

QY 61 SPARKYDQLVTRVVTHEMAHAGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120
 DB 61 SPARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 XX
 AC AAB36343;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.

PF 05-APR-2000; 200CWO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/7/TA) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 68.2%; Score 528; DB 3; Length 151;
Best Local Similarity 68.9%; Pred. No. 1,1e-45;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAATAIVFSGSALAGVVPQYGGGNGGNGNSGPNSEINITYQGGNSALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQGFNRNATIDQWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGCGNGADVGGSDSSIDLITQRFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
ABR82651
ID ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
OS Escherichia coli.
XX
PN WO2003064446-A2.
XX
PD 07-AUG-2003.
XX
PP 30-JAN-2003; 2003WO-EP000943.
XX
PR 31-JAN-2002; 2002GB-0002275.
XX
PA (HANS-) HANSA MEDICAL RES AB.
XX
PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-846136/61.
DR N-PSDB; ACF36153.
XX
PT New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
PS Disclosure; Page 41-42; 42pp; English.
XX
CC The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
SQ Sequence 151 AA;

Query Match 67.6%; Score 523; DB 7; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.4e-45;
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYQYGSANAALQ 60
Db 1 MKLLKVEATAIVFSGSALAGVVPQYGGGNGGNGNSGPNSEINITYQGGNSALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQGFNRNATIDQWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGCGNGADVGGSDSSIDLITQRFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

Search completed: August 2, 2004, 14:48:26
Job time : 45.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|----------------------|
| 1 | 684 | 88.4 | 151 | 1 | US-08-233-788A-59 |
| 2 | 504 | 65.1 | 120 | 1 | US-08-233-788A-57 |
| 3 | 83.5 | 10.8 | 208 | 4 | US-09-252-991A-27661 |
| 4 | 83.5 | 10.8 | 738 | 3 | US-08-864-038A-3 |
| 5 | 79.5 | 10.3 | 321 | 4 | US-09-498-520A-18 |
| 6 | 79.5 | 10.3 | 975 | 4 | US-09-328-352-4764 |
| 7 | 78 | 10.1 | 363 | 1 | US-08-458-023B-6 |
| 8 | 78 | 10.1 | 435 | 2 | US-08-331-515A-2 |
| 9 | 78 | 10.1 | 435 | 3 | US-09-168-406A-2 |
| 10 | 77.5 | 10.0 | 254 | 3 | US-09-128-450-26 |
| 11 | 77.5 | 10.0 | 254 | 4 | US-09-823-494-26 |
| 12 | 77.5 | 10.0 | 364 | 1 | US-07-792-259-17 |
| 13 | 77 | 9.9 | 943 | 4 | US-09-056-586-204 |
| 14 | 77 | 9.9 | 943 | 4 | US-09-072-596-199 |
| 15 | 77 | 9.9 | 943 | 4 | US-09-477-135A-131 |
| 16 | 77 | 9.9 | 943 | 4 | US-09-072-967-234 |
| 17 | 76.5 | 9.9 | 1739 | 4 | US-09-540-236-3739 |
| 18 | 76 | 9.8 | 252 | 4 | US-09-431-887-32 |
| 19 | 76 | 9.8 | 273 | 4 | US-09-328-352-6167 |
| 20 | 76 | 9.8 | 892 | 4 | US-09-336-447A-5 |
| 21 | 76 | 9.8 | 2123 | 3 | US-08-968-685A-10 |
| 22 | 75.5 | 9.8 | 232 | 4 | US-09-555-352-10 |
| 23 | 75.5 | 9.8 | 254 | 4 | US-09-431-887-23 |
| 24 | 75.5 | 9.8 | 415 | 4 | US-09-025-769B-280 |
| 25 | 75 | 9.7 | 208 | 3 | US-09-128-450-18 |
| 26 | 75 | 9.7 | 208 | 4 | US-09-823-494-18 |
| 27 | 75 | 9.7 | 349 | 4 | US-09-300-971A-9 |

| | | | | | |
|----|------|-----|-----|---|-------------------|
| 28 | 74.5 | 9.6 | 304 | 1 | US-07-851-976B-8 |
| 29 | 74.5 | 9.6 | 304 | 1 | US-08-291-609-8 |
| 30 | 74.5 | 9.6 | 304 | 1 | US-08-401-136-8 |
| 31 | 74.5 | 9.6 | 304 | 1 | US-08-850-554-8 |
| 32 | 73.5 | 9.5 | 211 | 1 | US-08-276-852-34 |
| 33 | 73.5 | 9.5 | 211 | 1 | US-08-133-011-16 |
| 34 | 73.5 | 9.5 | 211 | 1 | US-08-322-730A-16 |
| 35 | 73.5 | 9.5 | 211 | 1 | US-08-387-874-16 |
| 36 | 73.5 | 9.5 | 211 | 1 | US-08-899-575-34 |
| 37 | 73.5 | 9.5 | 211 | 1 | US-08-899-575-34 |
| 38 | 73.5 | 9.5 | 211 | 2 | US-08-383-619-16 |
| 39 | 73.5 | 9.5 | 211 | 3 | US-08-907-739-16 |
| 40 | 73.5 | 9.5 | 211 | 4 | US-09-729-597-16 |
| 41 | 73.5 | 9.5 | 211 | 5 | PCT-US93-08364-16 |
| 42 | 73.5 | 9.5 | 211 | 5 | PCT-US95-08743-34 |
| 43 | 73.5 | 9.5 | 238 | 4 | US-09-495-880A-42 |
| 44 | 73.5 | 9.5 | 266 | 4 | US-09-495-880A-42 |
| 45 | 73.5 | 9.5 | 293 | 3 | US-08-438-745-4 |

ALIGNMENTS

RESULT 1

US-08-233-788A-59

; Sequence 59, Application US/08233788A

; Patent No. 5635617

; GENERAL INFORMATION:

; APPLICANT: Doran, James L.

; APPLICANT: Kay, William W.

; APPLICANT: Clouthier, Sharon C.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

; TITLE OF INVENTION: OF SALMONELLA

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,788A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 682-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 59:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 151 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-233-788A-59

Query Match 88.4%; Score 684; DB 1; Length 151;

Best Local Similarity 90.1%; Pred. No. 5,5e-66;

Matches 136; Conservative 3; Mismatches 12; Indels 0;

Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSNAALAQ 60


```
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match      10.8%; Score 83.5; DB 3; Length 738;
Best Local Similarity 27.7%; Pred. No. 2;
Matches 44; Conservative 12; Mismatches 56; Indels 47; Gaps 6;

QY 3 LKVAAPRAIVVSSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSNAALALQSD 62
DB 419 LKRSASASASASASAG-----CGGGGGNGGNGGGGG-----GGAGALAAALAA 465
QY 63 ARKYDQL-----VTRVVTHEMAHAGQADNSTIELTQNGFRNNAIDOWNAKNSDITVGYQY 118
DB 466 AGAGGLGGGGGGGALAAALAAAGAGG-----GGFGGLGL-----GGL 504
QY 119 GGNNAALVNQASDSS-----VMVRQVGFNNATA 148
DB 505 CGGSAALAAAAAASGGGGRALRRALRRMRGGGSA 543

RESULT 5
US-09-498-520A-18
; Sequence 18, Application US/09498520A
; Patent No. 6613553
; GENERAL INFORMATION:
; APPLICANT: Rock, Charles O
; APPLICANT: Heath, Richard J
; TITLE OF INVENTION: No. 6613553al Enoyl Reductases and Methods of Use Thereof
; FILE REFERENCE: SJ-0022
; CURRENT APPLICATION NUMBER: US/09/498,520A
; CURRENT FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-498-520A-18

Query Match      10.3%; Score 79.5; DB 4; Length 321;
Best Local Similarity 24.5%; Pred. No. 1; 7;
Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;

QY 4 LKVAAPRAIVVSSA-----LAG-----VVPQGGGNGHNG----- 34
DB 109 LKAAGLKVMVCGAVKHAKAEQAQGDGAVICQGGGGGHTGLVGTPLVAQAVEAVKIPV 168
QY 35 ---CGNSSGPDSTLSIYQYGSNAALALQSDARKYDQVLTVRVTHEMAHAG-----Q 83
DB 169 VAAGGLHDG-----RGAA-AALALGAQG---VMGTFRFIASHBAHAGDLYRQAVE 215
QY 84 GADNSTIEL-TQNG-----FRNNATIDOWNAKNSDITV-----GOYGGNNAALVN 127
DB 216 AADEDVTVTRCYSGKPMVKKNPYVDWEARPGDIQPPQAMVSIIRGANGGIGGQIEG 275
QY 128 QTADSSVMVRQVGF 143
DB 276 LDKAKSCFAMGQSAGG 291

RESULT 6
US-09-328-352-4764
; Sequence 4764, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Berton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
```

```
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match      10.3%; Score 79.5; DB 4; Length 975;
Best Local Similarity 25.0%; Pred. No. 8;
Matches 34; Conservative 19; Mismatches 70; Indels 13; Gaps 4;

QY 15 SSGALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSNAALALQSDARKYDQVLTVRV 74
DB 279 NCTGDSG-VSALGGSGSGDGAGNGIASGNGEHNHYGNG---NGDDVDITAPITGYL 333
QY 75 THE-----MAHAGQADNSTIELTQNGFRNNAIDOWNAKNSDITVGYGGNNAALVNQ 128
DB 334 NISGNSFTLIGNSSSSSVNTAPTTSNTVNDNTID--NGNSGGTGGSGSGSGDGLNG 391
QY 129 TASDSSVMVRQVGF 144
DB 392 AASGNGEHNHYGNGN 407

RESULT 7
US-08-458-023B-6
; Sequence 6, Application US/08458023B
; Patent No. 5667990
; GENERAL INFORMATION:
; APPLICANT: Beika, Randy M.
; APPLICANT: Yoder, Wendy
; APPLICANT: Takagi, Shinobu
; APPLICANT: Boomathathan, Karuppan C.
; TITLE OF INVENTION: ASPERGILLUS EXPRESSION SYSTEM
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5667990o No. 5667990disk of No. 5667990th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,023B
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowrey Dr., Karen A.
; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 4086.010-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-023B-6

Query Match      10.1%; Score 78; DB 1; Length 363;
Best Local Similarity 30.7%; Pred. No. 2.9;
Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps 8;
```



```
RESULT 11
US-09-823-494-26
; Sequence 26, Application US/09823494
; Patent No. 6355610
; GENERAL INFORMATION:
; APPLICANT: Chesebro, Bruce W
; APPLICANT: Caughey, Byron W
; APPLICANT: Chabry, Joelle
; APPLICANT: Priola, Susette
; TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion
; FILE OF INVENTION: Protein
; FILE REFERENCE: 50121
; CURRENT APPLICATION NUMBER: US/09/823,494
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/128,450
; PRIOR FILING DATE: 1998-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 26
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hamster sp.
US-09-823-494-26

Query Match      10.0%; Score 77.5; DB 4; Length 254;
Best Local Similarity 25.4%; Pred. No. 2;
Matches 32; Conservative 19; Mismatches 19; Indels 15; Gaps 6;

QY 26 WG-CGGNHNGGNSGPDSTLSIYQGSANAALQSDARKYDQLVTRVTHEMAHAGQG 84
DB 89 WQGGGTHNQNKSKPTNNK--HMAGAAAAGAVVGLGY--MLGSAMSRPMHFGND 144
QY 85 ADNSTIELTQGFENNA---TIDQNAKNS-----DITVQYQGNNAAL-VNQTSADSS 134
DB 145 WEDRYRENVMRYNQVYRVDQYNNQNFVHDCVNITIKQHTVTTTGTGTFEDIK 204
QY 135 VMVRQV 140
DB 205 IMERVV 210

RESULT 12
US-07-792-259-17
; Sequence 17, Application US/07792259
; Patent No. 5286638
; GENERAL INFORMATION:
; APPLICANT: TANAKA, YOSHIKAZU
; APPLICANT: ASHIKARI, TOSHIHIKO
; APPLICANT: HATANAKA, HARUYO
; APPLICANT: SHIBANO, YUJI
; APPLICANT: AMACHI, TERUO
; APPLICANT: NAKAYAMA, TORU
; TITLE OF INVENTION: SUMIDA, MOTOO
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DABBY & CUSHMAN
; STREET: 1615 L. STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,259
; FILING DATE: 19911115
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.

Query Match      9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;
```

```
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 9437/93433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-792-259-17

Query Match      10.0%; Score 77.5; DB 1; Length 364;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 13; Mismatches 28; Indels 39; Gaps 8;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGHNH---GGNSSG-----PDSTLS 46
DB 1 MKLSLFSIFAIVIGALAL---PQPGGGGSGVTCGGQSTNSQCCVWFVDVLDLQTN 56
QY 47 IYQYGSANAALQSDARKYDQLVTRVTHE-----MAHAGQ-----GADNSTI 90
DB 57 FYQ-GS-----KCESPVRK---ILRIVFHAIGFSPALTAAGQFGGGGADGSI 101

RESULT 13
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-204

Query Match      9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;
```

TREAT

| | | | |
|----|-----|---|-----|
| Qy | 16 | GSALAGVVPQWCGG--GNEN--GGNSSGPDSITLSIYQYGSANAA--LALQSDARKYDOLVLT | 71 |
| Db | 464 | GSNGIGFVNVGSGS:GNTNIGSN-----LGYNIGFNGVDYVNVFGNAGDFNQGF | 516 |
| Qy | 72 | RVVTHEMAHAGQADNSTIETL-----QNGFRNATIDOWNAKNSDITVQYGGNNAALVNO | 128 |
| Db | 517 | NTGNNNIGPANTGNNNIGIGLSDGNQQGFN-----IASGMNSTGNSGLFNSGTTNVGIFNA | 573 |
| Qy | 129 | TASDSSVMVRQVGFGNNTAN | 149 |
| Db | 574 | GTGN-----VGIANSSTGN | 587 |

```

RESULT 14
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350

```

| | Query Match | 9.9%; | Score 77; | DB 4; | Length 943; |
|---|-----------------------|---|------------------|----------------|--------------------|
| | Best Local Similarity | 26.2%; | Pred. No. 14; | | |
| | Matches | 37; | Conservative 14; | Mismatches 66; | Indels 24; Gaps 7; |
| y | 16 | GSALGCVVPQWGGG-GNHN-GGNSSGPDSLTLSIYQGSANAA--LALQSDARKYDGLVLT | 71 | | |
| | | | | | |
| b | 464 | GSGNTGVFNVGSGSLGNYNGSGN-----LGTYNIGFNGVDYNVGFCGNAGDFNQGFA | 516 | | |
| | | | | | |
| y | 72 | RVVTHEMAHAGGAGDANSITIELT---QNGFRNATIDQWNAKNSDITTVGYQGNNAALVWQ | 128 | | |
| | | | | | |
| b | 517 | NTGNNNIGFATGNNGNIGIGSGDQQGFN---TASGWSGTGNSGLFNSGTNNVGTFNA | 573 | | |
| | | | | | |
| | 129 | TASDSSVMVRQVGFGNNTATN | 149 | | |

```

Db      574 GTGN-----VGIANSGTN 587
      :|||:|:|:|
      :|||:|:|:|

RESULT 15
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match          9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7

Qy      16 GSALAGVVPQKGG-GNHN-CGGNSGPDSTLTSTVQVGSANAA--LALQSDARKYDQLVT 71
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      571 GSGNIGVFNVSGSLGNTNIGSN-----LGIYNIQFNVDYVNGNAGDFNQGA 623
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      72 RYVTHEMAHAGQADNSTIELT---QNGFRNNATIDQWNAKNSDITVGYQGNNAALVNQ 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      624 NTGNNNIGFANTGNNGNIGLGLSGDNQQGFN---IASGWSGTGNSGLFNSGTTNNVGFNA 680
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      129 TASDSSVMVQVGFNNATAN 149
      :|||:|:|:|
Db      681 GTGN-----VGIANSGTN 694
      :|||:|:|:|

Search completed: August 2, 2004, 14:58:34
Job time : 13 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 525 | 67.8 | 151 | 12 | US-09-741-873B-4 |
| 2 | 525 | 67.8 | 151 | 12 | US-09-741-873B-4 |
| 3 | 447 | 57.8 | 131 | 12 | US-09-741-873B-2 |
| 4 | 447 | 57.8 | 131 | 12 | US-09-741-873B-2 |
| 5 | 101.5 | 13.1 | 445 | 15 | US-10-369-493-20638 |
| 6 | 94.5 | 12.2 | 438 | 14 | US-10-156-761-9343 |
| 7 | 91 | 11.8 | 186 | 12 | US-10-282-122A-49412 |
| 8 | 85 | 11.0 | 3705 | 12 | US-10-282-122A-77944 |
| 9 | 85 | 11.0 | 6310 | 12 | US-10-282-122A-67793 |
| 10 | 84.5 | 10.9 | 1778 | 14 | US-10-238-075-749 |
| 11 | 84 | 10.9 | 1129 | 12 | US-10-282-122A-8048 |
| 12 | 83.5 | 10.8 | 278 | 9 | US-09-810-264-28 |
| 13 | 83 | 10.7 | 597 | 9 | US-09-793-306-146 |
| 14 | 83 | 10.7 | 678 | 12 | US-10-282-122A-64573 |
| 15 | 82.5 | 10.7 | 435 | 14 | US-10-128-714-3213 |

| | | | | | | |
|----|------|------|------|----|----------------------|--------------------|
| 16 | 82.5 | 10.7 | 515 | 14 | US-10-128-714-8213 | Sequence 8213, Ap |
| 17 | 82.5 | 10.7 | 2732 | 14 | US-10-238-075-1119 | Sequence 1119, Ap |
| 18 | 82.5 | 10.7 | 2834 | 14 | US-10-085-959-252 | Sequence 252, App |
| 19 | 82 | 10.6 | 65 | 9 | US-09-996-194-16 | Sequence 16, Appl |
| 20 | 82 | 10.6 | 65 | 12 | US-10-164-966-33 | Sequence 33, Appl |
| 21 | 82 | 10.6 | 354 | 10 | US-09-820-843A-21 | Sequence 21, Appl |
| 22 | 82 | 10.6 | 440 | 12 | US-10-424-599-22922 | Sequence 22922, A |
| 23 | 82 | 10.6 | 1862 | 12 | US-10-282-122A-49757 | Sequence 49757, A |
| 24 | 81.5 | 10.5 | 2204 | 12 | US-10-282-122A-64364 | Sequence 64364, A |
| 25 | 81 | 10.5 | 562 | 12 | US-10-282-122A-54514 | Sequence 54514, A |
| 26 | 81 | 10.5 | 628 | 12 | US-10-282-122A-53269 | Sequence 53269, A |
| 27 | 80.5 | 10.4 | 154 | 16 | US-10-437-963-162284 | Sequence 162284, A |
| 28 | 80.5 | 10.4 | 486 | 15 | US-10-369-493-20619 | Sequence 20619, A |
| 29 | 80 | 10.3 | 145 | 16 | US-10-437-963-147748 | Sequence 147748, A |
| 30 | 80 | 10.3 | 477 | 12 | US-10-425-114-70098 | Sequence 70098, A |
| 31 | 80 | 10.3 | 507 | 12 | US-10-424-599-229226 | Sequence 229226, A |
| 32 | 80 | 10.3 | 538 | 12 | US-10-425-114-68152 | Sequence 68152, A |
| 33 | 80 | 10.3 | 558 | 16 | US-10-437-963-175203 | Sequence 175203, A |
| 34 | 79.5 | 10.3 | 562 | 14 | US-10-156-761-13039 | Sequence 13039, A |
| 35 | 79.5 | 10.3 | 1721 | 12 | US-10-282-122A-62548 | Sequence 62548, A |
| 36 | 79 | 10.2 | 209 | 12 | US-10-424-599-321110 | Sequence 221110, A |
| 37 | 79 | 10.2 | 276 | 15 | US-10-369-493-3641 | Sequence 3641, Ap |
| 38 | 79 | 10.2 | 688 | 14 | US-10-032-585-7876 | Sequence 7876, Ap |
| 39 | 79 | 10.2 | 974 | 12 | US-10-282-122A-44999 | Sequence 44999, A |
| 40 | 78.5 | 10.1 | 271 | 14 | US-10-156-761-11721 | Sequence 11721, A |
| 41 | 78.5 | 10.1 | 292 | 16 | US-10-437-963-195404 | Sequence 195404, A |
| 42 | 78.5 | 10.1 | 472 | 16 | US-10-467-479-2 | Sequence 2, Appli |
| 43 | 78 | 10.1 | 545 | 15 | US-10-369-493-18473 | Sequence 18473, A |
| 44 | 78 | 10.1 | 594 | 14 | US-10-156-761-13173 | Sequence 13173, A |
| 45 | 78 | 10.1 | 1346 | 12 | US-10-282-122A-49773 | Sequence 49773, A |

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 67.8%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 66-46; Indels 0;
Matches 103; Conservative 21; Mismatches 27; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIVQYGSANAALAQ 60

```

Db      1  MLLKVAALAVFGSAGVVPQYGGGNGHGGGNSGPNSENIYYOYGGNSALALQ 60
QY      61  SPARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db      61  TDARNSDLTITQHGCGGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNQTASDSSVMVROVFGNNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQFGNNATAHQY 151

```

RESULT 2

```

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match      67.8%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 6e-46;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1  MLLKVAALAVFGSAGVVPQYGGGNGHGGGNSGPDSTLSIYOGSANAALALQ 60
Db      1  MLLKVAALAVFGSAGVVPQYGGGNGHGGGNSGPNSENIYYOYGGNSALALQ 60
QY      61  SPARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db      61  TDARNSDLTITQHGCGGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNQTASDSSVMVROVFGNNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQFGNNATAHQY 151

```

RESULT 3

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      57.8%; Score 447; DB 12; Length 131;
Best Local Similarity 64.9%; Pred. No. 5.2e-38;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY      21  GVVPQYGGGNGHGGGNSGPDSTLSIYOGSANAALALQSDARKYDQLVTRVVTHEMAH 80
Db      1  GVVPQYGGGNGHGGGNSGPNSENIYYOYGGNSALALQTDARNSDLTITQHGCGGAD 60
QY      81  AGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGGNNAALVNQTASDSSVMVROV 140
Db      61  VGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQ 120
QY      141  GFGNNATAHQY 151
Db      121  GFGNNATAHQY 131

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

Query Match

57.8%; Score 447; DB 12; Length 131;

Best Local Similarity 64.9%; Pred. No. 5.2e-38;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

Qy 21 GVVPQWGGGHNHGGSSGPDSTLSIYQGSANAALQSDARKYDOLVTRVVTHMAH 80
Db 1 GVVPQYGGGHNHGGSSGNSPNSLTYQYGGNSALALQTDARNSLTITQHGCGNGAD 60

Qy 81 AGQAGDNSTIELTQNGFRNNATIDWNKNSDITVGYGNNALVNOTASDSVMYRQV 140
Db 61 VQGSDDSSIDLTRGFGNSATLQWNGKNSMTVKQFGGNGAAVQDTASNSVNVTVQ 120

Qy 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.1%; Score 101.5; DB 15; Length 445;
Best Local Similarity 26.0%; Pred. No. 0.076;
Matches 45; Conservative 20; Mismatches 55; Indels 53; Gaps 7;

Qy 7 AAFAA-----IVVSGSALAGVVPQWGGG-----NHNG-----GNSSGPDSTLSIYQY 50
Db 19 AAFADGNTVYLNQTDQANITQSGNNSVGFNGSGFLQENGTLGA-NLLTVKQS 77

Qy 51 GSANAALQSDARKYDOLVTRVVTHMAHAGGA-----DNSTIELTQNGFRNNATID 104
Db 78 GNSNS-----VGRDIQKQSGAGNSAALFQGTGSDVLELQQTGTSNGAVPS 123

Qy 105 QWNAKN-----SPITVGYGNNALVNOTASDSVMYRQV 141
Db 124 GWNWINDPGVFNKITQDSSNSGSKSVIQLGKNVFSIKQNTGNSVNVQIG 176

RESULT 6
US-10-156-761-9343
; Sequence 9343, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, NASHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9343
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9343

Query Match 12.2%; Score 94.5; DB 14; Length 438;
Best Local Similarity 25.9%; Pred. No. 0.39;
Matches 37; Conservative 27; Mismatches 46; Indels 33; Gaps 8;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGHNHGGSSGPDSTLSIYQGSANAALA-- 58
Db 1 MRSIRAAAVGAVTMSLALAAASAC---GGSGTGGGSDSP-KTLT-YWASNGGASIAVD 54

Qy 59 ---LQSDARKYDQ-----LVTRVWTHMAHAGGADNSTIELTQN-GFRN 99
Db 55 KXVLPELDKFEQGIKVKLEWVPNSDLLNRLT--ATTSGGPDVNLNIGNTWSASLOA 112

Qy 100 NATIDWNKNSDITVGYGNN 122
Db 113 TGALLPWDARFED---KIGGKD 131

RESULT 7
US-10-282-122A-49412
; Sequence 49412, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carx, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16


```

Qy 10 AAIIVVGSALAGVVPQW-----GGGG-----NHNGGNSG---PDSTLSIYQGSAN 54
Db 111 AGITAGSAM-----PTWAIAGLSLLGVGAAADNGGGSGGSDPDTs-----156
Qy 55 AALALQSDARKYDOLVTRVVTWTHEMAHAGQAGADNSTIEL-----TQNGFRNN 100
Db 157 -----APATPIDLLVS---PDGLRLTGRGEAGTTVIRDAAGNLIGSGIVGADGNFVT 207
Qy 101 ATIDQWAKNSDIIVQYGGNNALVNQTSADSSVMVRQVGFNNATAN 149
Db 208 LNAQPINSENLDVTLTDAGNVSPAGVATPAPDATAPLATPDLAINEQCN 256

RESULT 10
US-10-238-075-749
; Sequence 749, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; TITLE OF INVENTION: E. coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 749
; LENGTH: 1778
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-749

Query Match 10.9%; Score 84.5; DB 14; Length 1778;
Best Local Similarity 26.8%; Pred. No. 25;
Matches 40; Conservative 21; Mismatches 59; Indels 29; Gaps 7;

Qy 5 KVAFAAIVVSGSALAGVVPQW-----GGGNNHNGGNSGPDSTLSIYQGSANAALAL- 59
Db 906 KISSNSTDAINGSLQYGVADSFSTYLGCGADISDTGVLSGPTVTIGTDTYNVGDALAI 965
Qy 60 -----QSDARKYDQLVTRVVTWTHEMAHAGQAGADNSTIELTQ--NGFRNNATIDQWNAKN 110
Db 966 NTFSTFSLGDLALWD-----ATAGKFSKHXGINNAPSIVITDVANGAVSSTSDAINGSQ 1019
Qy 111 ----SDITVQYGGNNALVNQTSADSSV 135
Db 1020 LYGVSDVIADALGN--AVVN---TDGSI 1043

RESULT 11
US-10-282-122A-48048
; Sequence 48048, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Oalsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

```



```
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3213
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3213
```

```
Query Match      10.7%; Score 82.5; DB 14; Length 435;
Best Local Similarity 19.6%; Pred.No.6.6;
Matches 31; Conservative 30; Mismatches 56; Indels 31; Gaps 6;

Qy 13 VVSGSALAGVVPQGGGNNHGGNSSGPDSTLSIYOYGSANAALALQSDARKYDQ--- 69
Db 274 VLTGHKGSVTCVRNGGTGKI-----YTSSTHDETIKIWNAQNGSLLQTLSAHAHRVNHLS 329

Qy 70 ---VTRVVTHEMAHAGQ--GADNSTIELTQNGFR-----NNATIDQ-----WN-107
Db 330 TDFALRTAYHD--HTGKVPGSDTEKVAVKRFEQAAMVNNKIVEKLVSASDDFTMYLWD 387

Qy 108 AKNSDITVGOYGGNNAALVNOTASDSSVMVYQVGFNN 145
Db 388 PENSTKPIARLLGHCKEVNHTVFPDPMAYIASAGFDNH 425
```

```
Search completed: August 2, 2004, 15:36:11
Job time : 36.8 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877,809 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
|------------|-------|-------------|--------|----|-------------|

| | | | | | | |
|----|-------|-------|------|----|---------------------|-------------------|
| 1 | 774 | 100.0 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 2 | 709 | 91.6 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 3 | 689 | 89.0 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 4 | 684 | 88.4 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 5 | 662 | 85.5 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 6 | 622 | 80.4 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 7 | 619 | 80.0 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 8 | 611 | 78.9 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 9 | 609 | 78.7 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 10 | 606 | 78.3 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 11 | 602 | 77.8 | 151 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 12 | 597 | 77.1 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 13 | 574 | 74.2 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 528 | 68.2 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 525 | 67.8 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 525 | 67.8 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 523 | 67.6 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 523 | 67.6 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 504 | 65.1 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 470 | 60.7 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 21 | 466 | 60.2 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 466 | 60.2 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 23 | 466 | 60.2 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 24 | 447 | 57.8 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 447 | 57.8 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 343 | 44.3 | 109 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 27 | 263 | 34.0 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 28 | 166.5 | 21.5 | 70 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 29 | 153 | 19.8 | 48 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 30 | 105.5 | 13.6 | 145 | 21 | US-09-739-449-8854 | Sequence 8854, Ap |
| 31 | 105.5 | 13.6 | 145 | 23 | US-09-803-110-8854 | Sequence 8854, Ap |
| 32 | 101.5 | 13.1 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 33 | 101.5 | 13.1 | 445 | 33 | US-60-360-039-20638 | Sequence 20638, A |
| 34 | 99 | 12.8 | 492 | 21 | US-09-708-427-7657 | Sequence 7657, Ap |
| 35 | 99 | 12.8 | 544 | 21 | US-09-708-427-7656 | Sequence 7656, Ap |
| 36 | 99 | 12.8 | 573 | 21 | US-09-708-427-7655 | Sequence 7655, Ap |
| 37 | 99 | 12.8 | 955 | 24 | US-09-935-625-7946 | Sequence 7946, Ap |
| 38 | 99 | 12.8 | 955 | 24 | US-09-935-625-25158 | Sequence 25158, A |
| 39 | 99 | 12.8 | 1036 | 24 | US-09-935-625-7945 | Sequence 7945, Ap |
| 40 | 99 | 12.8 | 1036 | 24 | US-09-935-625-25157 | Sequence 25157, A |
| 41 | 99 | 12.8 | 1055 | 19 | US-09-570-581A-1851 | Sequence 1851, Ap |
| 42 | 99 | 12.8 | 1055 | 19 | US-09-573-655A-349 | Sequence 349, App |
| 43 | 99 | 12.8 | 1055 | 19 | US-09-573-655A-667 | Sequence 667, App |
| 44 | 99 | 12.8 | 1055 | 19 | US-09-573-655B-349 | Sequence 349, App |
| 45 | 99 | 12.8 | 1055 | 19 | US-09-573-655B-667 | Sequence 667, App |

ALIGNMENTS

RESULT 1
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-20

| | | | | |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match | 100.0% | Score 774; | DB 19; | Length 151; |
| Best Local Similarity | 100.0%; | Pred. No. 3.3e-75; | | |
| Matches 151; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | MKLLKVAFAAIIVSGSALAGVFPQWGGGNGHNGGNSGPDSTLSIYYGSAANAALQ | 60 | |
| Db | 1 | MKLLKVAFAAIIVSGSALAGVFPQWGGGNGHNGGNSGPDSTLSIYYGSAANAALQ | 60 | |
| Qy | 61 | SDARKYDQLVTRVVTHEVAHAGQAGADNSTIELTQNGFRNNAIDOWNAKNSDITVYGQGG | 120 | |
| Db | 61 | SDARKYDQLVTRVVTHEVAHAGQAGADNSTIELTQNGFRNNAIDOWNAKNSDITVYGQGG | 120 | |

RESULT 2

```

US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteridis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

```

```

Query Match      91.6%; Score 709; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 3.8e-68;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAFAAIIVVGSGALAGVVPQGGGGNHNGGNSGGPDSTLSIIYQGSANAALQ 60
    |||||
Db 1 MKLLKVAFAAIIVVGSGALAGVVPQGGGGNHNGGNSGGPDSTLSIIYQGSANAAL--- 57
    |||||

QY 61 SDARKYDQLVTRVVTTHMAHA-----GQAGDNSTIELTQGFNNATIDQWNAKNSD 112
    |||||
Db 58 ----YDQLVTRVVTTHMAHAGYGVQAGADNSTIELTQGFNNATIDQWNAKNSD 112
    |||||

QY 113 ITVGQYGGNNAAALVNQTASDSSVVMRVQVGFNNATANQY 151
    |||||
Db 113 ITVGQYGGNNAAALVNQTASDSSVVMRVQVGFNNATANQY 151
    |||||

```

RESULT. T 3

```

RESULTS 3
US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407

```

```

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PR1
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

```

| | | | | |
|---------------------------|--------|---|-----------|-------------|
| Query Match | 89.0%; | Score 689; | DB 19; | Length 151; |
| Best Local Similarity | 90.7%; | Pred. No. 5.6e-66; | | |
| Matches 137; Conservative | 3; | Mismatches 11; | Indels 0; | Gaps 0; |
| Qy | 1 | MKLLKVAFAAIVVSGSALAGVPQWGGGNGHNGGNSGPGDSTLSIYGYGSANAALAQ | 60 | |
| Db | 1 | MKLLKVAFAAIVVSGSALAGVPQWGGGNGHNGGNSGPGDSTLSIYGYGSANAALAQ | 60 | |
| Qy | 61 | SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDWNAKNSDIITVGXCG | 120 | |
| Db | 61 | SDARKSEITIQTSGVGCNCGADVQGQADNSTIELTQNGFRNNATIDWNAKNSDIITVGXCG | 120 | |
| Qy | 121 | NNAALVNQTASDSSVMVRQVGFGNNTANQY | 151 | |
| Db | 121 | NNAALVNQTASDSSVMVRQVGFGNNTANQY | 151 | |

RESULT 4

```

US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

```

Query Match 88.4%; Score 684; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.9e-65;
Matches 136: Conservative 3; Mismatches 12; Indels 0; Gaps 0;

```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGAGDNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNPAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 5
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match 85.5%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-63; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQ 105
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----FRNNATIDQ 105
QY 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 80.4%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2.2e-58; Indels 44; Gaps 2;
Matches 129; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGFR 98
Db 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHA----- 103
QY 99 NNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 104 -----DOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 80.4%; Score 622; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1e-58; Indels 46; Gaps 2;
Matches 128; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIELTQNGF 97
Db 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADYQGGADNSTIELTQNGF 97
QY 98 NNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 98 RNNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 80.0%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2.2e-58; Indels 44; Gaps 2;
Matches 129; Conservative 0; Mismatches 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGFR 98
Db 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHA----- 103
QY 99 NNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 104 -----DOWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match      78.9%; Score 611; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-57;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      78.7%; Score 609; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 2.7e-57;
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
```

```
QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151
Db 121 LVTRVVTHEMAHASVVMVQVGFNNATANQY 151

RESULT 10
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      78.3%; Score 606; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 5.7e-57;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120
QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151
Db 121 HEMAHANQTASDSSVVMVQVGFNNATANQY 151

RESULT 11
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-31

Query Match      77.8%; Score 602; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.3e-56;
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKYDQLVTRVVTHEMAH 80
```


Db 1 GVFPWGGGNGHNGSGDPDSTLSIYQGSANAALALOSDARKSETTTTQSGYNGAD 60
QY 81 AGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQV 140
Db 61 VQCADNSTIETQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQV 120

QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 77.1%; Score 597; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQCADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.2%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.7e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGGDPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQCADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.2%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGNPNSELNIIYQGGNSALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 TDARNSDLTITQHGNGGADVGQSDSDSIDLTQRFNGNSATLDQWNGKNSMTVKQFG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATANQY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

Tue Aug 3 10:54:37 2004

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      67.8%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.5e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIYQYGSANAALALQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 MKLLKVAIAIAIVFSSAVAGVVPQYGGGNNHGGSSGPNSEINLYQYGGNSALALQ 60
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      61 SPARKYDQLVTRVTHEVAHAGQAGADNSTIELTQNGFENNATIDOWNAKNSDITVQYGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db      61 TDARNSDLTITQHGGNGADVGQGSDDSSIDLQRFNGSATLDQWNGKNSMTVKQFGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      121 NNAALVNQTSASSVMVRQVGFNNATANQY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 GNGRAVDQTASNSVNVTVQVGFNNATAHQY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: August 2, 2004, 15:26:44
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-20
Perfect score: 774
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 525 | 67.8 | 151 | 5 | US-09-741-873C-4 |
| 2 | 447 | 57.8 | 131 | 5 | US-09-741-873C-2 |
| 3 | 87 | 11.2 | 573 | 7 | US-60-565-632-7907 |
| 4 | 87 | 11.2 | 573 | 7 | US-60-579-062-7907 |
| 5 | 85 | 11.0 | 147 | 6 | US-10-425-115-193207 |
| 6 | 85 | 11.0 | 412 | 7 | US-60-565-632-7905 |
| 7 | 85 | 11.0 | 412 | 7 | US-60-579-062-7905 |
| 8 | 85 | 11.0 | 841 | 7 | US-60-565-632-7906 |
| 9 | 85 | 11.0 | 841 | 7 | US-60-579-062-7906 |
| 10 | 81.5 | 10.5 | 436 | 7 | US-60-581-351-7636 |
| 11 | 80.5 | 10.4 | 511 | 6 | US-10-425-115-320950 |
| 12 | 80 | 10.3 | 581 | 1 | PCT-US04-07412-1695 |
| 13 | 80 | 10.3 | 581 | 6 | US-10-389-559-1695 |
| 14 | 79.5 | 10.3 | 321 | 7 | US-60-556-841-11319 |
| 15 | 79 | 10.2 | 374 | 1 | PCT-US04-11210-37 |
| 16 | 79 | 10.2 | 719 | 5 | US-09-248-796A-17559 |
| 17 | 79 | 10.2 | 719 | 5 | US-09-248-796A-17559 |
| 18 | 78.5 | 10.1 | 753 | 6 | US-10-170-205E-35514 |
| 19 | 78.5 | 10.1 | 753 | 6 | US-10-170-205E-35514 |
| 20 | 77 | 9.9 | 545 | 7 | US-60-565-632-7973 |
| 21 | 77 | 9.9 | 545 | 7 | US-60-579-062-7973 |
| 22 | 76.5 | 9.9 | 179 | 6 | US-10-767-701-35342 |
| 23 | 76.5 | 9.9 | 254 | 1 | PCT-US04-16242-10 |
| 24 | 76 | 9.8 | 556 | 6 | US-10-425-115-337674 |
| 25 | 76 | 9.8 | 892 | 5 | US-09-952-267B-5 |
| 26 | 76 | 9.8 | 892 | 6 | US-10-872-768-5 |

US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION: Staffan
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,978
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

ALIGNMENTS

RESULT 1

Query Match 67.8%; Score 525; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 8.3e-40;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | MLLKVAFAAIVVSGSALAAGVVPQGGGNGGNGGSSGDPSTLSIYQVGSANAALQ | 60 |
| DB | 1 | MLLKVAFAAIVVSSAVAGVVPQGGGNGGNGGNGGNSGPNSELNIYQVGGNSALQ | 60 |
| QY | 61 | SDARKYDQLVTRVYVTHEMAHAGQADNSTIELTQNGFRNATIDQWNAKNSDITVQCYGG | 120 |
| DB | 61 | TDARNESDLTITQHGCGGADVGQSDSSIDLTLQRFNGSATLDQWNGKNSMTVQKFGG | 120 |
| QY | 121 | NNAALVNQATSSDSSVMVRQVGFNNATANY | 151 |
| DB | 121 | GNAAVDQATSSNVVTVQVGFNNATAHQY | 151 |

Sequence 5, Appli
Sequence 273, App
Sequence 123, App
Sequence 242035,
Sequence 339993,
Sequence 280, App
Sequence 16, Appl
Sequence 38, Appl
Sequence 11867, A
Sequence 25337, A
Sequence 346132,
Sequence 8923, Ap
Sequence 9813, Ap
Sequence 553, App
Sequence 343835,
Sequence 22578, A
Sequence 9640, Ap
Sequence 13, Appl

```
RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8601723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

Query Match 57.8%; Score 447; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 7,1e-33;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVVPWGGGNGHNGGSGPDSTLSIYQGSANAALQSDARKYDQLVTRVVTHEMAH 80
Db 1 GVVPWGGGNGHNGGSGPNSELTYQGGNSALALQTDARNSLTITQHGGGNGAD 60
QY 81 AGQAGDNSTIELTQGFNNATIDOWNAKNSDITVQYGGNNAALVNOTASDSVVVRQV 140
Db 61 VQGGSDSSIDITQGFNSATLDOWNGKNSMTVQFGGNGAADVQDTASNSSVNTQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 3
US-60-565-632-7907
; Sequence 7907, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7907
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Diabrotica virgifera

US-09-543-407-20.rapn

Query Match 11.2%; Score 87; DB 7; Length 573;
Best Local Similarity 24.9%; Pred. No. 9.5;
Matches 42; Conservative 15; Mismatches 62; Indels 50; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 241 GNENGTGAENNAADAQTDVAQ--GSTNEAENNAADVQNDAAQENGAAGAENSNGAD 299
QY 81 AGQAGDN-STIELTQN-----GFRNNATIDOWN----- 107
Db 300 AAQGTGNGAAAEENTGNADPAQGNNDNGAAENGNGENGTAAENNAADVQNDAAQVNDNGA 359
QY 108 --AKNSDITVQYGGNNAALVNOTASD-----SSVMVRQVGFNNATAN 149
Db 360 AAENNGNADAAQSDNDNGAAAEENTTNADAQNGAAQSTANEANAENNAAD 408

RESULT 4
US-60-579-062-7907
; Sequence 7907, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21(53403)C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7907
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
US-60-579-062-7907

Query Match 11.2%; Score 87; DB 7; Length 573;
Best Local Similarity 24.9%; Pred. No. 9.5;
Matches 42; Conservative 15; Mismatches 62; Indels 50; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
Db 241 GNENGTGAENNAADAQTDVAQ--GSTNEAENNAADVQNDAAQENGAAGAENSNGAD 299
QY 81 AGQAGDN-STIELTQN-----GFRNNATIDOWN----- 107
Db 300 AAQGTGNGAAAEENTGNADPAQGNNDNGAAENGNGENGTAAENNAADVQNDAAQVNDNGA 359
QY 108 --AKNSDITVQYGGNNAALVNOTASD-----SSVMVRQVGFNNATAN 149
Db 360 AAENNGNADAAQSDNDNGAAAEENTTNADAQNGAAQSTANEANAENNAAD 408

RESULT 5
US-10-425-115-193207
; Sequence 193207, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
```

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 193207
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MET4577_107790C.1.pep
US-10-425-115-193207

Query Match 11.0%; Score 85; DB 6; Length 147;
Best Local Similarity 24.6%; Pred. No. 2.8;
Matches 35; Conservative 18; Mismatches 63; Indels 26; Gaps 4;

QY 7 AAPAIVVSGSALAGVVPONGGGNHGNGSGPDSLTSLIYQY-----GSANAALALQSD 62
DB 27 AQIPASATTGAATPAPQPGSGTGGVGVGAPDAPLELYMHDIILGSGSPRA----- 80
QY 63 ARKYDQLVTRVVTTHMAGAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYQYGN 122
DB 81 -----RPIITGLGNIYNGVPPFAPRPIGFSAPRNGVAIPNANGQ-----VPTYNGNT 126
QY 123 AALVNQTASDSSVMVRQVGFN 144
DB 127 GIPLDTGLSRAGFL--QPTG 146

RESULT 6

US-60-565-632-7905
; Sequence 7905, Application US/60565632

; GENERAL INFORMATION:

; APPLICANT: Monsanto Technology, LLC

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K.

; APPLICANT: Larosa, Thomas J.

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K.

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; TITLE OF INVENTION: Compositions Thereof

; FILE REFERENCE: 38-21(53403)B

; CURRENT APPLICATION NUMBER: US/60/565,632

; CURRENT FILING DATE: 2004-04-27

; NUMBER OF SEQ ID NOS: 15449

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7905

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Diabrotica virgifera

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (369)..(369)

; OTHER INFORMATION: Xaa can be any naturally occurring amino acid

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (392)..(392)

; OTHER INFORMATION: Xaa can be any naturally occurring amino acid

US-60-565-632-7905

Query Match 11.0%; Score 85; DB 7; Length 412;

Best Local Similarity 26.4%; Pred. No. 9.6;

Matches 43; Conservative 12; Mismatches 62; Indels 46; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQYGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80

DB 85 GNENGTTGAENNAADAQTDAQ--GSTNEAENNAADVQNDAAQANENGAAENGSGNAD 143

QY 81 AQCGADN-STIELTON-----GFRNATIDOWN----- 107

DB 144 AAQGTDNAAAEENTGNADPAQGNNGAAENGSGNENGTAENNAADVQNDAAQVNNNGA 203

QY 108 --AKNSDITVGYGGNNAALVNQTASDSSVMVRQVGFGNATA 148
DB 204 AAENGNADAQSNNDNGAAENNTNADAQNDAAQ--GTANEANA 245

RESULT 7

US-60-579-062-7905

; Sequence 7905, Application US/60579062

; GENERAL INFORMATION:

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K

; APPLICANT: Larosa, Thomas J

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; TITLE OF INVENTION: Compositions thereof

; FILE REFERENCE: 38-21 (53403) C

; CURRENT APPLICATION NUMBER: US/60/579,062

; CURRENT FILING DATE: 2004-06-11

; NUMBER OF SEQ ID NOS: 41445

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7905

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Diabrotica virgifera

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (369)..(369)

; OTHER INFORMATION: Xaa can be any naturally occurring amino acid

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (392)..(392)

; OTHER INFORMATION: Xaa can be any naturally occurring amino acid

US-60-579-062-7905

Query Match 11.0%; Score 85; DB 7; Length 412;

Best Local Similarity 26.4%; Pred. No. 9.6;

Matches 43; Conservative 12; Mismatches 62; Indels 46; Gaps 7;

QY 30 GNHNGGG--NSSGPDSTLSIYQYGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80

DB 85 GNENGTTGAENNAADAQTDAQ--GSTNEAENNAADVQNDAAQANENGAAENGSGNAD 143

QY 81 AQCGADN-STIELTON-----GFRNATIDOWN----- 107

DB 144 AAQGTDNAAAEENTGNADPAQGNNGAAENGSGNENGTAENNAADVQNDAAQVNNNGA 203

QY 108 --AKNSDITVGYGGNNAALVNQTASDSSVMVRQVGFGNATA 148

DB 204 AAENGNADAQSNNDNGAAENNTNADAQNDAAQ--GTANEANA 245

RESULT 8

US-60-565-632-7906

; Sequence 7906, Application US/60565632

; GENERAL INFORMATION:

; APPLICANT: Monsanto Technology, LLC

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K.

; APPLICANT: Larosa, Thomas J.

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K.

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; TITLE OF INVENTION: Compositions Thereof

; FILE REFERENCE: 38-21(53403)B

; CURRENT APPLICATION NUMBER: US/60/565,632

; CURRENT FILING DATE: 2004-04-27

```
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7906

Query Match          11.0%; Score 85; DB 7; Length 841;
Best Local Similarity 30.7%; Pred. No. 23;
Matches 42; Conservative 13; Mismatches 56; Indels 26; Gaps 8;

QY 30 GNHNGGG--NSSGPPSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GNENGTGAENNAADAQTDAQAQ--GSTNEAENNAADVQNDAAQACANENGAAENSGNAD 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 81 AQCGADN-STIELTON-----GFRNATIDQWNAKNSDITVQYQGNNAALVNOTASDSS 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 AAQGTIDNGAAAGTGNADPAQNDNGAA-----AENSGNENGTAAGNNA---NPDVQND 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 135 VMVROVGF--NNATAN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 AQVNDNGTAAENGNAD 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-60-579-062-7906
; Sequence 7906, Application US/60579062
; GENERAL INFORMATION:
; APPLICANT: Baum, James A
; APPLICANT: Kovacic, David K
; APPLICANT: Larosa, Thomas J
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: 38-21 (53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7906
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-579-062-7906

Query Match          11.0%; Score 85; DB 7; Length 841;
Best Local Similarity 30.7%; Pred. No. 23;
Matches 42; Conservative 13; Mismatches 56; Indels 26; Gaps 8;

QY 30 GNHNGGG--NSSGPPSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GNENGTGAENNAADAQTDAQAQ--GSTNEAENNAADVQNDAAQACANENGAAENSGNAD 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 81 AQCGADN-STIELTON-----GFRNATIDQWNAKNSDITVQYQGNNAALVNOTASDSS 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 AAQGTIDNGAAAGTGNADPAQNDNGAA-----AENSGNENGTAAGNNA---NPDVQND 379
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 135 VMVROVGF--NNATAN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 AQVNDNGTAAENGNAD 396
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-60-579-902-7377
; Sequence 7377, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; FILE REFERENCE: 38-21 (53444) A
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 7377
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-7377

Query Match          11.0%; Score 85; DB 7; Length 2663;
Best Local Similarity 24.7%; Pred. No. 91;
Matches 42; Conservative 23; Mismatches 59; Indels 46; Gaps 7;

QY 5 KVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQSDAR 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 QLLAKSLVVSASAI-----DNROKITSSGGDITLTGQDNLGKIA---GQ 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 65 KYDQLVTRVVTHEMAHAGQCADNSTIELTQNGFRNA-----TIDQWNAKNS- 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 HHQLQNTQSIINKEGH--MSADTVDTINTHQGLNNTAGLIVAERNMILRTGELLNRQGSV 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 112 ----DITVQYQY-----GNNALVNOTASDSSVMRVQVFGNNATAN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 RSGQDLTLNTHGNQLDNRDSGLQGLFSQGA-----MHLNTGYLNNQSGH 518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-60-581-351-7636
; Sequence 7636, Application US/60581351
; GENERAL INFORMATION:
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants
; FILE REFERENCE: 38-21 (53372) B
; CURRENT APPLICATION NUMBER: US/60/581,351
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: US 60/479,962
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 13980
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7636
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Ceratopteris richardii
US-60-581-351-7636

Query Match          10.5%; Score 81.5; DB 7; Length 436;
Best Local Similarity 26.1%; Pred. No. 21;
Matches 29; Conservative 16; Mismatches 43; Indels 23; Gaps 3;

QY 29 GGNHNGGNS---SGPDSITLSIYQGSANAALALQSDARKYDQLVTRVVTHEMAHAGOGA 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 77 GAGAGGGGSSVAGNAGTANNINQOHSAAEAGLLAIARS-----GGDLAQSGGG 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 86 DNSTIELTQNGFRNATIDQWNAKNSDITVQYQGNNAALVNOTASDSSVM 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 RGNLLDLHSD-----TANSSDLVEGHEGGHGGHNOQDSQVL 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-10-425-115-320950
; Sequence 320950, Application US/10425115
```

```
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants
/ FILE REFERENCE: 38-2153222B
/ CURRENT APPLICATION NUMBER: US/10/425,115
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 369326
/ SEQ ID NO 320950
/ LENGTH: 511
/ TYPE: PRT
/ ORGANISM: Zea mays
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(511)
/ OTHER INFORMATION: unsure at all Xaa locations
/ FEATURE:
/ OTHER INFORMATION: Clone ID: MRT4577_55769C.1.pap
US-10-425-115-320950

Query Match          10.4%; Score 80.5; DB 6; Length 511;
Best Local Similarity 20.8%; Pred. No. 32;
Matches 38; Conservative 28; Mismatches 70; Indels 47; Gaps 6;

QY 1 M K L I K V A P A A I V S -----G S A L A G V V P Q W G G -----G N H N G G N S S G P D 42
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D b 1 3 3 L R L L E A Q E A I L A S C R D V L R A S G F R F D A W A K V I P G S D E G Y A W A A N Y A L G R L G G D P N 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 4 3 S T L S I Y Q Y G S A N A L A L O S D A -----R K Y D Q L V T R V V T H E M A H A G Q A D N S T I E - 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D b 1 9 3 K T V G I I E G S A Q A L T F V S D E V L P P K L S Y N T F G E T T Y L T N S F L N F G Q N A A Q D S H E M 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 9 2 - L T Q N G F R N A T I D -----Q M A K N S D I T V G Y G G N N A A L V N O T A S D S S V W V R Q V G F G N N A 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D b 2 5 3 L R S R G S F X N G T L A D P C A P R G Y S R N E E M L R M S G A S R S T L E N Q -----Y V N A G 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 T A N 149
D b 300 T G N 302

RESULT 13
PCT-US04-07412-1695
/ Sequence 1695, Application PC/TUS0407412
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Wang, Jian-rui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yunging
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Goodrich, Ryle W.
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Weng, Gezhi
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Boyle, Bryan J.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 822CIP/PCT
/ CURRENT APPLICATION NUMBER: PCT/US04/07412
/ CURRENT FILING DATE: 2004-03-19
/ PRIOR APPLICATION NUMBER: US 10/389,559
/ PRIOR FILING DATE: 2003-03-14
/ PRIOR APPLICATION NUMBER: US 60/365,264
```

```
/ PRIOR FILING DATE: 2002-03-14
/ PRIOR APPLICATION NUMBER: US 60/340,187
/ PRIOR FILING DATE: 2001-12-12
/ PRIOR APPLICATION NUMBER: US 10/296,115
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: PCT/US00/35017
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: US 09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ PRIOR APPLICATION NUMBER: US 09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ PRIOR APPLICATION NUMBER: US 10/275,027
/ PRIOR FILING DATE: 2001-01-25
/ PRIOR APPLICATION NUMBER: PCT/US01/02623
/ PRIOR FILING DATE: 2001-01-25
/ PRIOR APPLICATION NUMBER: US 09/491,404
/ PRIOR FILING DATE: 2000-01-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1920
/ SOFTWARE: dt_FL_genes Version 6.0
/ SEQ ID NO 1695
/ LENGTH: 581
/ TYPE: PRT
/ ORGANISM: Homo sapiens
PCT-US04-07412-1695

Query Match          10.3%; Score 80; DB 1; Length 581;
Best Local Similarity 22.3%; Pred. No. 41;
Matches 27; Conservative 25; Mismatches 47; Indels 22; Gaps 5;

QY 4 5 L S I Y Q Y - G S A N A L A L Q S D A R K Y D Q L - - V T R V V T H E M A H A G Q A D N S T I E L T Q N G F R N N A 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D b 24 2 V S L Y P L A G G A T Q A F A X E N N C K A Y K E T Y G V S H I T R H D M L Q I P K Q Q N E K Y Q V P Q - - - F D Q S 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 10 2 T I -----D O W N A K N S D I T V G Y G G N N A A L V - - - N O T A S D S S V W V R Q V G F G N N 145
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D b 2 9 9 T I R N I E S A K G L D V M D S W P L Q N A D G T V A E Y N G V H V F A L A G S P K D A D D T S I Y M F Y K V G D N 358
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 14 6 A 146
D b 3 5 9 S 359

RESULT 14
US-10-389-559-1695
/ Sequence 1695, Application US/10389559
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Wang, Jian-rui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yunging
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Goodrich, Ryle W.
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Weng, Gezhi
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Boyle, Bryan J.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 822
/ CURRENT APPLICATION NUMBER: US/10/389,559
/ CURRENT FILING DATE: 2003-03-14
/ PRIOR APPLICATION NUMBER: US 60/365,264
/ PRIOR FILING DATE: 2002-03-14
/ PRIOR APPLICATION NUMBER: US 60/340,187
/ PRIOR FILING DATE: 2001-12-12
```

```
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1920
; SOFTWARE: pt.fl_genes Version 6.0
; SEQ ID NO 1695
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-389-559-1695

Query Match      10.3%; Score 80; DB 6; Length 581;
Best Local Similarity 22.3%; Pred. No. 41;
Matches 27; Conservative 25; Mismatches 47; Indels 22; Gaps 5;

QY 45 LSIYQY-GSANAALQSDARKYDQL--VTRVVTHEMAHAGQADNSTIELTONGFRNNA 101
Db 242 VSLYPLAGGATCAFKENNQKAYKETGYVSHTRHMLQIPKQQQNEKYQVPQ---FDOS 298

QY 102 TI-----DOWNAKNSDITVQYGGNNAALV-----NOTASDSSVMVRQVGFNN 145
Db 299 TIKNIESAKGLDWSNPLQADGTVAEYNGYHVVFALAGSPKADDTISYMFYKVGDN 358

QY 146 A 146
Db 359 S 359

RESULT 15
US-60-556-841-11319
; Sequence 11319, Application US/60556841
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)
; CURRENT APPLICATION NUMBER: US/60/556,841
; CURRENT FILING DATE: 2004-03-25
; NUMBER OF SEQ ID NOS: 12463
; SEQ ID NO 11319
; LENGTH: 321
; TYPE: PRT
; ORGANISM: Caulobacter crescentus CB15
; US-60-556-841-11319

Query Match      10.3%; Score 79.5; DB 7; Length 321;
Best Local Similarity 24.5%; Pred. No. 22;
Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;

QY 4 LKVAAPAAATVWSGSA-----LAG---VVPQWGGGNNHG----- 34
Db 109 LKAAGLKNVAVGAVKHAVKACQAGCDAVICQGGGGGHTGLVTLPLVAQAVEAVKIPV 168

QY 35 ---GNNSSGPDSTLSYQYGSANAALQSDARKYDQLVTRVVTHEMAHAG-----Q 83
Db 169 VAAGGLHDG-----RGLA-AALALGAQG---VVMGTRFIASHEAHAGDLVRQAVVE 215

QY 84 GADNSTIEL-TONG---FRNNATIDOWNAKNSDITV-----GQYGGNNAALVN 127
Db 216 AADEDTVTRCYSGKPMRVKKNPYVDDWEARPDIOPTPQQAAMVSIIRNGAMGGIGGQIEG 275
```

```
QY 128 QTASDSSVMVRQVGF 143
Db 276 LDKAKSCFAMQSQSAGG 291
```

Search completed: August 2, 2004, 15:29:53
Job time : 18.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAAPAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 689 | 89.0 | 151 | 2 JC6039 | fimbrin protein ag |
| 2 | 689 | 89.0 | 151 | 2 A10635 | major curlin chain |
| 3 | 528 | 68.2 | 151 | 2 S70788 | curlin protein csg |
| 4 | 506.5 | 65.4 | 152 | 2 D90806 | curlin major subun |
| 5 | 506.5 | 65.4 | 152 | 2 H85665 | hypothetical prote |
| 6 | 105.5 | 13.6 | 145 | 2 A31143 | conserved hypothet |
| 7 | 105.5 | 13.6 | 145 | 2 H98144 | hypothetical prote |
| 8 | 99 | 12.8 | 573 | 2 C86266 | F319.21 protein - |
| 9 | 96.5 | 12.5 | 151 | 2 S70787 | curlin nucleator p |
| 10 | 96.5 | 12.5 | 151 | 2 C90806 | minor curlin subun |
| 11 | 96.5 | 12.5 | 151 | 2 G85665 | curlin minor chain |
| 12 | 96 | 12.4 | 2174 | 2 E35965 | hypothetical glyci |
| 13 | 92.5 | 12.0 | 1567 | 2 S11672 | ice nucleation pro |
| 14 | 91.5 | 11.8 | 1322 | 2 S07053 | ice nucleation pro |
| 15 | 91 | 11.8 | 602 | 1 P10221 | leishmanolysin (EC |
| 16 | 91 | 11.8 | 645 | 2 F70825 | probable PPE prote |
| 17 | 90 | 11.6 | 639 | 2 C42049 | leishmanolysin (EC |
| 18 | 89.5 | 11.6 | 374 | 2 T03875 | probable homeobox |
| 19 | 89.5 | 11.6 | 375 | 2 T03874 | probable homeobox |
| 20 | 89.5 | 11.6 | 447 | 2 G84687 | probable disease r |
| 21 | 89 | 11.5 | 652 | 2 E97857 | cell surface antig |
| 22 | 88.5 | 11.4 | 151 | 2 JC6040 | fimbrin protein ag |
| 23 | 88.5 | 11.4 | 151 | 2 AH0635 | nucleation compone |
| 24 | 88.5 | 11.4 | 1588 | 2 A96036 | probable adhesin Z |
| 25 | 88.5 | 11.4 | 1588 | 2 H91188 | probable adhesin E |
| 26 | 88 | 11.4 | 590 | 1 A45621 | leishmanolysin (EC |
| 27 | 88 | 11.4 | 599 | 2 B42049 | leishmanolysin (EC |
| 28 | 88 | 11.4 | 599 | 2 A44951 | leishmanolysin (EC |
| 29 | 88 | 11.4 | 646 | 1 S19916 | leishmanolysin (EC |

30 87.5 11.3 575 2 S35327 protein kinase egg
31 87 11.2 1748 2 S42136 cniB protein - Tet
32 86 11.1 1635 2 A10452 hemolysin [importe
33 85.5 11.0 438 2 T35789 probable secreted
34 85 11.0 615 2 E70663 probable PPE prote
35 85 11.0 891 2 E96590 hypothetical prote
36 85 11.0 3705 2 AD0123 probable autotrans
37 84.5 10.9 188 1 YQBCF2 F7-2 fimbrial prot
38 84.5 10.9 1034 2 JC2143 ice nucleation act
39 84.5 10.9 1258 2 QJ0188 ice nucleation pro
40 84 10.9 494 2 T15502 hypothetical prote
41 83 10.7 407 2 F87494 hypothetical prote
42 83 10.7 409 2 T20847 probable PPE prote
43 83 10.7 590 2 E70946 probable PPE prote
44 83 10.7 678 2 A70762 CRBB-binding prote
45 83 10.7 3190 2 T13828

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C;Species: Salmonella enteritidis
C;Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C;Accession: JC6039; PC6015; A44898
R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bauser, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A;Title: Salmonella enteritidis agfAC operon encoding thin, aggregative fimbriae.
A;Reference number: JC6039; MUID:96146512; PMID:8550497
A;Accession: JC6039
A;Molecule type: DNA
A;Residues: 1-151 <COL>
A;Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43599.1; PID:gl184714
A;Accession: PC6015
A;Molecule type: protein
A;Residues: 21-52 <CO2>
A;Experimental source: strain 27855-3b
A;Note: the authors translated the codon ACG for residue 44 as Ile
R;Collinson, S.K.; Emdy, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A;Title: Purification and characterization of thin, aggregative fimbriae from Salmonell
A;Reference number: A44898; MUID:91310586; PMID:1677357
A;Contents: 27855
A;Accession: A44898
A;Status: preliminary
A;Molecule type: protein
A;Residues: 21-33 <CO3>
A;Note: sequence extracted from NCBI backbone (NCBIP:45936)
C;Genetics:
A;Gene: agfa
C;Function:
A;Description: major component of thin aggregative fimbriae
A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C;Keywords: fimbria
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 89.0%; Score 689; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.1e-51;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNNHNGGNSGSPDSTLSIYQVGSANAALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNNHNGGNSGSPDSTLSIYQVGSANAALQ 60
QY 61 SDARKYDQLVTRVVTTHMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNTASDSSVMVRQVGFNNATANYQ 151

Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, M.; Mouton, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 89.0%; Score 689; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.1e-51;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQVLRVVTTHMAHAGQADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
DB 61 SPARKSETITQSGYGVNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:gl147558; PIDN:CAA62282.1; PID:gl147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
A:.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AB000205; GB:U00096; NID:gl1787265; PIDN:AAC74126.1; PID:gl1787279;
A:Experimental source: strain K-12, substrain M6165
R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The Ppos sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6,'V', 8-151 <OLSI>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42;44-50 <OLS2>
R:Olsen, A.N.; Arngvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133,'RQRSGWUW' <OLS3>
A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:G290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 58.2%; Score 528; DB 2; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-37;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQVLRVVTTHMAHAGQADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
DB 61 TDARNSDLTITQHGCGNGADVQGGSDSDITLQTFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 4

D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain R1MD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11238796
A:Accession: D90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:gl3360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECS1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.7e-36;
Matches 102; Conservative 21; Mismatches 26; Indels 1; Gaps 1;
QY 1 MKLLKVAAPAAIVSGSALAGVVPQW-GGGGNGHNGSGSPDSTLSIYQYGSANAALAL 59
DB 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAL 60
QY 60 QSDARKYDQVLRVVTTHMAHAGQADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 119

QY 3 LLKVAFAAIVVSSALAGVVPPQGG-----GNNHNGGNSSGPDPSTLSIYYQGSANA 55
 :
 Db 1 MIRKSFTASALVALVLGLSAAAPAMANDVRIEQYGWNSAGGAQEYGNGRIRTYQNGGYNR 60
 :
 QY 56 ALALQSARKYDQLVTRVVTWTHMAHAGQGADNSTIELTQTGFRRNATIDQNAKNSDITV 115
 :
 Db 61 IVGHQ-----YGR-----HNLSAVQGEHDNYGSTTQGNRVVAGI----- 96
 :
 QY 116 GQYGGNNAALVNQTASDVSVVMVRVQVGFNNTAQ 150
 :
 Db 97 GQFGSNHTTLTDGNGNIAAGVQVGRGCSANVSQ 131
 :

RESULT 7

H98144
 hypochemical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C)
 C|Species: Agrobacterium tumefaciens
 C|Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C|Accession: H98144
 R|Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Ma
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Ma
 Science 294, 2323-2328, 2001
 A|Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobact
 A|Reference number: A97359; MUID:21608551; PMID:11743194
 A|Accession: H98144
 A|Status: preliminary
 A|Molecule type: DNA
 A|Residues: 1-145 <KUR>
 A|Cross-references: GB:AEO007870; PIDN:AAK88682.1; PID:g15158413; GSPDB:GN00170
 C|Genetics:
 A|Gene: AGR_L_228
 A|Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
 Best Local Similarity 23.2%; Pred. No. 0.049;
 Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY 3 LLKVAFAAIVVSSALAGVVPPQGG-----GNNHNGGNSSGPDPSTLSIYYQGSANA 55
 :
 Db 1 MIRKSFTASALVALVLGLSAAAPAMANDVRIEQYGWNSAGGAQEYGNGRIRTYQNGGYNR 60
 :
 QY 56 ALALQSARKYDQLVTRVVTWTHMAHAGQGADNSTIELTQTGFRRNATIDQNAKNSDITV 115
 :
 Db 61 IVGHQ-----YGR-----HNLSAVQGEHDNYGSTTQGNRVVAGI----- 96
 :
 QY 116 GQYGGNNAALVNQTASDVSVVMVRVQVGFNNTAQ 150
 :
 Db 97 GQFGSNHTTLTDGNGNIAAGVQVGRGCSANVSQ 131
 :

RESULT 8

C86266
 F3F19.21 protein - Arabidopsis thaliana
 C|Species: Arabidopsis thaliana (mouse-ear cress)
 C|Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
 C|Accession: C86266
 R|Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; D
 ansen, N.F.; Hughes, B.; Huizlar, L.
 Nature 408, 816-820, 2000
 A|Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.
 C.A.; Li, J.H.; Li, Y.Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.;
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A|Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; H
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A|Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A|Reference number: A86141; MUID:21016719; PMID:11130712
 A|Accession: C86266
 A|Status: preliminary
 A|Molecule type: DNA
 A|Residues: 1-573 <STO>
 A|Cross-references: GB:AE0051172; NID:g4850402; PIDN:AAD31072.1; GSPDB:GN00141

A;Map position: 1

Query Match 12.8%; Score 99; DB 2; Length 573;
 Best Local Similarity 25.4%; Pred. No. 0.81; Mismatches 24; Conservative 34; Gaps 6;
 Matches 34; Conservative 24; Mismatches 24; Indels 30; Gaps 6;
 QY 13 VVSGALAGVWPQWGGGNGHNGSGPDTLSIYQGSANAALALQSDARKYDQVTR 72
 DB 84 IVSGGTVEG---KYRNDGGHNG---ISGPDTRSDVYPQASSFGAKGLNID----- 127
 QY 73 VVTHEMAHAGQADNSTIELTQNGFRNNA-TIDOWNAKNS-----DITVGOYGGNN 122
 DB 128 IQSNKIAQGG---STTVLNNHGFSGNAVNVPMPVHNSYGAPPQGAQIIPVQMSVNP 183
 QY 123 AALVNOTASDSSVM 136
 DB 184 NVMMKSPQSPVV 197

RESULT 9

S70787
 curlin nucleator protein csGB precursor - Escherichia coli (strain K-12)
 N;Alternate names: csGB protein; curlin nucleation component; minor curlin protein
 C;Species: Escherichia coli
 C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
 C;Accession: S70787; F64846
 R;Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
 Mol. Microbiol. 18, 661-670, 1995
 A;Title: Expression of two csG operons is required for production of fibronectin- and Cc
 A;Reference number: S70783; MUID:96414468; PMID:8817489
 A;Accession: S70787
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-151 <HAM>
 A;Cross-references: EMBL:X90754; NID:G1147558; PIDN:CAA62281.1; PID:G1147563
 A;Experimental source: strain K12, substrain W3110
 A;Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A;Title: The complete genome sequence of Escherichia coli K-12.
 A;Reference number: A64720; MUID:97426617; PMID:9278503
 A;Accession: F64846
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-151 <BLAT>
 A;Cross-references: GB:AE000205; GB:U00096; NID:G1787265; PIDN:AAC74125.1; PID:G1787278;
 A;Experimental source: strain K-12, substrain MG1655
 C;Genetics:
 A;Gene: csGB

A;Map position: 23.15

C;Function:
 A;Description: minor component of wild-type curl; interaction between CsgA and CsgB tri
 A;Note: curl are thin, coiled fibers expressed on the surface of Escherichia coli that
 and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-151/Product: minor curlin chain #status predicted <MAP>

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29; Mismatches 15; Indels 5; Gaps 2;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDTLSIYQGSANAALALQSDARKYDQVTRVVTHEMAHAGQADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 77 SNRAKIDQGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 10

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subst
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: C90806
 R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 R;Sawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 GSAwara, N., 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: C90806
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA34842.1; PID:G13360879; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RIMD 0509952
 C;Genetics:
 A;Gene: ECs1419

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29; Mismatches 15; Indels 5; Gaps 2;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDTLSIYQGSANAALALQSDARKYDQVTRVVTHEMAHAGQADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 77 SNRAKIDQGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 11

G85665
 curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: G85665
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: G85665
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-151 <STO>
 A;Cross-references: GB:AE005174; NID:G12514573; PIDN:AAGS5787.1; GSPDB:GN0145; UWGP:Z16
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: csGB

Query Match 12.5%; Score 96.5; DB 2; Length 151;
 Best Local Similarity 28.1%; Pred. No. 0.29; Mismatches 15; Indels 5; Gaps 2;
 Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDTLSIYQGSANAALALQSDARKYDQVTRVVTHEMAHAGQADNSTIELTQNGF 97
 DB 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNNSAQLRQGGSKLLAVVAQEGS 76
 QY 98 RNNATIDOWNAKNSDITVGOYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151
 DB 77 SNRAKIDQGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 12

E95965
 hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C;Species: Sinorhizobium meliloti
 C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C;Accession: E95965
 R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo

A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E95965
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2174 <KUR>
A;Cross-references: GB:AL591985; PIDN:CA49389.1; PID:g15140875; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymb
R;Gailbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: Smb21548
A;Genome: plasmid

Query Match 12.4%; Score 96; DB 2; Length 2174;
Best Local Similarity 27.0%; Pred. No. 6.5;
Matches 40; Conservative 18; Mismatches 54; Indels 36; Gaps 7;
QY 11 AIVVSGSALAGVWPO--WGGGNGHNGGNSGPDSTLSIYOYGS-----ANAA----- 56
DB 693 AIAATAGAGVILIAQSIGGGG--GGNATGGAGRGSGFQGGGGGGYANTANYGFK 749
QY 57 -LALQSDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDIT 115
DB 750 -GLTLTTQGSAAAGIVAQSV-----GGGGGTGTASSYSAGIGFTAS-----VAV 793
QY 116 QYGGNNAA--LVNQTSADSSVMVRQVG 141
DB 794 GGTGNGGAGGEVSVLSLTSAIRTCGGG 821

RESULT 13
S11672
ice nucleation protein - Xanthomonas campestris
C;Species: Xanthomonas campestris
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
C;Accession: S11672
R;Zhao, J.; Orser, C.S.
Mol. Gen. Genet. 223, 163-166, 1990
A;Title: Conserved repetition in the ice nucleation gene *inx* from Xanthomonas campestris
A;Reference number: S11672; MUID:91080859; PMID:2259339
A;Accession: S11672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1567 <ZHA>
A;Cross-references: EMBL:X52970; NID:g48531; PIDN:CAA37140.1; PID:g48532
C;Superfamily: ice nucleation protein

Query Match 12.0%; Score 92.5; DB 2; Length 1567;
Best Local Similarity 25.8%; Pred. No. 8.9;
Matches 46; Conservative 27; Mismatches 56; Indels 49; Gaps 11;
QY 14 VSGSALAG-----VVPQWGG--GGNHN-----GGGNSGPDSTLSIYOYGSANAALAL 59
DB 205 VYGSTLTADQSRVLVAGVGTETAGDHSDDLIAVGSTGTAGSDSI-LAGVSTQTAAGR 263
QY 60 QSDARKYDQLVT-----RVVTHEMAHAGQADNSTIELTQNGFRNNATI----- 103
DB 264 STLTAGYGTQTAQGSRLTSGYGTATSGSDSAVI-----SGYGTQTAGSESSLTAGY 319
QY 104 -DOWNAKNSDITVQYGG-----GNNALV-----NOTASDSSVMVRQVGFNNATNQ 150
DB 320 STQTARKGSDITAG-YGSTGTAGSDSALIAGVGTQTAGSESSLT--AGVGTQTARK 374

RESULT 14
S07053

ice nucleation protein *inaA* - Erwinia ananas
C;Species: Erwinia ananas
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999
C;Accession: S07053
R;Abe, K.; Watabe, S.; Emori, Y.; Watanabe, M.; Arai, S.
FEBS Lett. 258, 297-300, 1989
A;Title: An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those of *Erwinia caroliniana* and *Erwinia ictioerythrae*.
A;Reference number: S07053; MUID:90092494; PMID:2599095
A;Accession: S07053
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1322 <AB>
A;Cross-references: GB:X17316; NID:g296095; PIDN:CAA35194.1; PID:g296096
C;Superfamily: ice nucleation protein

Query Match 11.8%; Score 91.5; DB 2; Length 1322;
Best Local Similarity 29.1%; Pred. No. 8.9;
Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;
QY 34 GCGNSGPDSTLSIYOYGSANAALALQSDARKYDQLVTRVVTHEMAHAGQADNSTIELT 93
DB 933 GSTSTAGPSSL-IAGYGSTQTA-----GNSILT-----AGYS-----T 967
QY 94 QNGFRNNATIDQWNAKNSDITVQYG-----GNNALV-----NOTASDSSVMVRQVGF 143
DB 968 QTG-----QENSLLTTG-YGSTTAGVSSLIAGVGTQTATSPKSTLM--AGYG 1013
QY 144 NNATANO 150
DB 1014 SSQTARE 1020

RESULT 15

PL0221

leishmanolysin (EC 3.4.24.36) precursor [validated] - Leishmania major
N;Alternate names: promastigote surface proteinase; surface endopeptidase glycoprotein
C;Species: Leishmania major
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 15-Sep-2000
C;Accession: PL0221; A27598; A60648
R;Button, L.L.; McMaster, W.R.
J. Exp. Med. 171, 589, 1990
A;Reference number: PL0221
A;Contents: erratum
A;Accession: PL0221
A;Molecule type: DNA
A;Residues: 1-602 <BUT>
A;Cross-references: GB:Y00647; NID:g9554; PIDN:CAA68673.1; PID:g9555
A;Note: this is a revision to the sequence from reference A27598
R;Button, L.L.; McMaster, W.R.
J. Exp. Med. 167, 724-729, 1988
A;Title: Molecular cloning of the major surface antigen of Leishmania.
A;Reference number: A27598; MUID:88154764; PMID:3346625
A;Accession: A27598

A;Status: significant sequence differences

A;Molecule type: DNA

R;Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.

Mol. Biochem. Parasitol. 37, 235-246, 1989

A;Title: Characterization of the promastigote surface protease of Leishmania as a membr

A;Reference number: A60648; MUID:90114330; PMID:2608099

A;Accession: A60648

A;Molecule type: protein

A;Residues: 101, E', 103-118, SV', 121-123 <BOU>

A;Experimental source: strain LEM513

R;Schlagenhauf, E.; Etges, R.; Metcalf, P.

submitted to the Brookhaven Protein Data Bank, March 1997

A;Reference number: A68135; PDB:1LML

A;Contents: annotation; X-ray crystallography, 1.86 angstroms, residues 100-407, 412-498

A;Note: strain JRC-L119

C;Complex: homodimer

C;Function:

A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residu

A;Note: the activated form can activate the proenzyme form

C;Superfamily: leishmanolysin

C:Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; lip
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-100/Domain: activation peptide #status predicted <ATP>
F:101-577/Product: leishmanolysin #status experimental <MAI>
F:578-602/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F:48,264,268,334/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:100-101/Cleavage site: Val-Val (autolytic) #status experimental
F:125-142,191-230,314-386,393-455,406-425,415-489,466-510,515-565,535-558/Disulfide bond
F:264,268,334/Binding site: zinc, catalytic (His) (active) #status experimental
F:265/Active site: Glu #status predicted
F:300/407/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:577/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form)

Query Match 11.8%; Score 91; DB 1; Length 602;
Best Local Similarity 37.5%; Pred. No. 4;
Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;

| | | | |
|----|-----|--|-----|
| QY | 63 | ARKYDQVTRVVTHEMAHA-----GGADNSTIELTQNGFRNNATIDQWAKNSDITV--- | 115 |
| Db | 251 | ASRYDQVTRVVTHEMAHALGSGPFEDARIV-----ANVNVVRGKNFDVVPVINS | 301 |
| QY | 116 | -----GOYG | 119 |
| Db | 302 | STAVAKAREQYG | 313 |

Search completed: August 2, 2004, 14:56:23
Job time : 9.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 689 | 89.0 | 151 | 1 | CSGA_SALTY |
| 2 | 528 | 68.2 | 151 | 1 | CSGA_ECOLI |
| 3 | 506.5 | 65.4 | 152 | 1 | CSGA_ECO57 |
| 4 | 96.5 | 12.5 | 151 | 1 | CSGB_ECOLI |
| 5 | 92.5 | 12.0 | 1567 | 1 | ICEN_XANCT |
| 6 | 91.5 | 11.8 | 1322 | 1 | ICEN_PANAN |
| 7 | 91 | 11.8 | 602 | 1 | GP63_LEIMA |
| 8 | 88.5 | 11.4 | 151 | 1 | CSGB_SALTY |
| 9 | 88.5 | 11.4 | 151 | 1 | GP63_LEIDO |
| 10 | 88 | 11.4 | 590 | 1 | GP63_LEICH |
| 11 | 88 | 11.4 | 599 | 1 | GP63_LEICH |
| 12 | 88 | 11.4 | 646 | 1 | GP63_LEIME |
| 13 | 86.5 | 11.2 | 369 | 1 | PST3_MYCAV |
| 14 | 84.5 | 10.9 | 188 | 1 | FMF2_ECOLI |
| 15 | 84.5 | 10.9 | 1034 | 1 | ICEN_PANAN |
| 16 | 84.5 | 10.9 | 1258 | 1 | ICEN_ERWHE |
| 17 | 83 | 10.7 | 678 | 1 | YF48_MYCTU |
| 18 | 82.5 | 10.7 | 592 | 1 | CSA_CITFR |
| 19 | 82.5 | 10.7 | 641 | 1 | IMD_ARTGO |
| 20 | 82.5 | 10.7 | 1028 | 1 | OVO_DROME |
| 21 | 81.5 | 10.5 | 392 | 1 | HME1_HUMAN |
| 22 | 81.5 | 10.5 | 1317 | 1 | N145_YEAST |
| 23 | 81 | 10.5 | 1656 | 1 | QXEB_RICJA |
| 24 | 80.5 | 10.4 | 1210 | 1 | ICEN_PSEPL |
| 25 | 78.5 | 10.1 | 472 | 1 | SACB_BACAM |
| 26 | 78.5 | 10.1 | 487 | 1 | Y442_MYCTU |
| 27 | 78 | 10.1 | 147 | 1 | HFAA_CAUCR |
| 28 | 78 | 10.1 | 363 | 1 | PER_COPCI |
| 29 | 77.5 | 10.0 | 254 | 1 | PRIO_MESAU |
| 30 | 77.5 | 10.0 | 364 | 1 | PER_ARTRA |
| 31 | 77.5 | 10.0 | 499 | 1 | YFZ5_METTF |
| 32 | 77 | 9.9 | 370 | 1 | PST3_MYCTU |
| 33 | 76 | 9.8 | 252 | 1 | PRIO_RABIT |

34 76 9.8 504 1 FLIC_SALMC Q06981 salmonella
35 76 9.8 955 1 FRU_DROME Q81R81 drosophila
36 76 9.8 1115 1 TBC2_CHLRE Q8VXP3 chlamydomon
37 75.5 9.8 254 1 PRIO_CIGR Q60506 cricetus
38 75.5 9.8 576 1 CEA7_ECOLI Q47112 escherichia
39 75.5 9.8 1196 1 ICEV_PSEEX Q34479 pseudomonas
40 75 9.7 172 1 CH18_DROME Q07184 drosophila
41 75 9.7 389 1 ACUC_STAAM Q99TC9 staphylococ
42 74.5 9.6 163 1 HCV_NATPH Q39442 natronomona
43 74.5 9.6 370 1 PGLI_PENOL Q9V834 penicillium
44 74.5 9.6 663 1 DUS8_MOUSE O09112 mus musculus
45 74.5 9.6 760 1 YBIL_ECOLI P73780 escherichia

ALIGNMENTS

RESULT 1
CSGA_SALTY STANDARD; PRT; 151 AA.
ID AC P52225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM144 OR STM181 OR STM176.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen K., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
Baker S., Basham D., Brooks K.M., Dowd L., White N., Farrar J.,
Cronin A., Davis P., Davies R., Chillingworth T., Connor P.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22511367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
Burland V., Kocoyanni V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
and CT18.";
RT J. Bacteriol. 185:2330-2337(2003).
RN [5]
RN SEQUENCE FROM N.A.
RP SPECIES=enteritidis; STRAIN=27655-3B;
RC MEDLINE=96146512; PubMed=8550497;
RX Collinson S.K., Cloughier S.C., Doran J.L., Baner P.A., Kay W.W.;
RA "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RT J. Bacteriol. 178:662-667(1996).
RN [6]
RN SEQUENCE OF 21-151 FROM N.A.
RP SPECIES=enteritidis; STRAIN=27655-3B;
RC MEDLINE=94013373; PubMed=8104955;
RX Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RA "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
RT the structural gene for thin, aggregative fimbriae.";
RT J. Clin. Microbiol. 31:2263-2273(1993).
RN [7]
RN SEQUENCE OF 21-33.
RP SPECIES=enteritidis; STRAIN=27655-3B;
RC MEDLINE=91310586; PubMed=1677357;
RX Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
RA "Purification and characterization of thin, aggregative fimbriae from
RT *Salmonella enteritidis*.";
RT J. Bacteriol. 173:4773-4781(1991).
RN [8]
RN FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF002301; CAA05317.1; -;
DR EMBL; AE008749; AAL20074.1; -;
DR EMBL; AL627269; CAD08268.1; -;
DR EMBL; AE016840; AAC09399.1; -;
DR EMBL; U43280; AAC43599.1; -;
DR FIC; JC6039; JC6039.
DR Stydane; SGI0608; csgA.
KW SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SVMVRQVGFNNATANY -> DSVTVQAS (IN
FT REF. 6)
SQ SEQUENCE 151 AA; 15305 MW; 87DADCD168621359 CRC64;
Query Match 89.0%; Score 689; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.9e-53;
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAALVVGSGALAGVFPQWGGGNGHNGSSGPDSTLSIYQVGSAAALQ 60
DY 1 MKLLKVAFAALVVGSGALAGVFPQWGGGNGHNGSSGPDSTLSIYQVGSAAALQ 60
QY 61 SDARKYDQLVTRVVTTHMAHAGQAGADNSTIELTQGFNNATIDQWNAKNSDITVGYGG 120
DY 61 SDARKSETTITQSGYNGADVQAGADNSTIELTQGFNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151
DY 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151

RESULT 2
CSGA_ECOLI STANDARD; PRT; 151 AA.
ID _CSGA_ECOLI
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
RT repression of csgA, the subunit gene of fibronectin-binding curli in
RT Escherichia coli.";
RT Mol. Microbiol. 7:523-536(1993).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
RT K-12.";
RT Mol. Microbiol. 18:661-670(1995).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RT Science 277:1453-1474(1997).
RN [4]
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Tkemoto K., Inada T., Itoh T., Kajiura M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RT DNA Res. 3:137-155(1996).
RN [5]
RN SEQUENCE OF 21-40.
RP STRAIN=K12 / YMEI;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
RT fibronectin binding in *Escherichia coli* HB101.";
RT Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RN SEQUENCE OF 21-31.
RP MEDLINE=91310586; PubMed=1677357;
RX Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
RA "Purification and characterization of thin, aggregative fimbriae from
RT *Salmonella enteritidis*.";
RT J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.


```
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L04979; AAA23616.1; -.
CC DR EMBL: X90754; CAM62282.1; -.
CC DR EMBL: AAC74126.1; -.
CC DR EMBL: D90741; BAA35832.1; -.
CC DR EMBL: D90742; BAA35840.1; -.
CC DR PIR: S70788; S70788.
CC ECoGene: EG11489; csGA.
CC FimBria; Signal; Complete proteome.
CC FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
CC FT CHAIN 21 151 A -> E (IN REF. 1).
CC FT CONFLICT 7
CC SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
Query Match 68.2%; Score 528; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 3.6e-39;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGQGADNSTIELTQNGFRNNATIDQWNAKSDITVGYQG 120
Db 61 TDARNSDLTITQHGCGGADVQGSDDSSIDLTQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
ID CSGA ECO57
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii X., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tode T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF275733; AAK53212.1; -.
CC DR EMBL: AE005315; AAG55788.1; -.
CC DR EMBL: AF002554; BAB34843.1; -.
CC DR PIR: D90806; D90806.
CC DR PIR: H85665; H85665.
CC FimBria; Signal; Complete proteome.
CC FT SIGNAL 1 20 BY SIMILARITY.
CC FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC FT CONFLICT 152 AA; 15099 MW; EE2D294DDE91243 CRC64;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GGGGNGHNGSGSPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAL 60
QY 60 QSDARKYDQLVTRVVTHEMAHAGQGADNSTIELTQNGFRNNATIDQWNAKSDITVGYQG 119
Db 61 QADARNSDLTITQHGCGGADVQGSDDSSIDLTQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 120 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 152
RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR E1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
```



```
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;
Query Match 12.0%; Score 92.5; DB 1; Length 1567;
Best Local Similarity 25.8%; Pred. No. 3.1;
Matches 46; Conservative 27; Mismatches 56; Indels 49; Gaps 11;

QY 14 VSGALAG-----VPOWGG---CGNHN-----GGNSSGPDSTLSIYQYGSNAALAL 59
DQ 205 VYGTSLTGADQSLRVAGYGTETAGDHSDLIAGYGTGTAGSDSSI-LAGYGSTQTAAGR 263
QY 60 QSDARKYDQVLT-----RVVTHEMAHAGOGADNKTIELTONGFRFNATI----- 103
DQ 264 STLTAGYGTGTAGSGRLTSGYGTSGSDSAVI-----SGYGTGTAGSESSLTAYG 319
QY 104 -DOWNAKNSDITVQYQ-----GNNALV-----NOTASDSSVMVRQVGFNNATANO 150
DQ 320 STQARKGSDITAG-YGSTGTAGSDSALIAGYGTGTAGSESSLT--AGYGSTQTARK 374

RESULT 6
ID ICEA PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein Inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity";
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17316; CAA35194.1; .
CC DR PIR; S07053; S07053.
CC DR HSSP; P06620; 1INA.
CC DR InterPro; IPR000258; Ice_nucleatn.
CC DR Pfam; PF00818; Ice_nucleation; 69.
CC DR PRINTS; PR00327; ICENUCLEATN.
CC DR PROSITE; PS00314; ICE_NUCLEATION; 49.
CC DR ICD; ICD000000000; Ice_nucleation; Repeat; Outer membrane.
CC FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1322 AA; 131094 MW; 89B0BE24AA837039 CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 1322;
Best Local Similarity 29.1%; Pred. No. 3.1;
Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;

QY 34 GGNSSGPDSTLSIYQYGSNAALALQSDARKYDQVLTVRVTHEMAHAGQADNSTIELT 93
```

```
Db 933 GSTAGPDDSL-IAGYGSTQTA-----GYNILT-----AGYGS-----T 967
QY 94 QNGFRNATIDOWNAKNSDITVQYQ-----GNNALV-----NOTASDSSVMVRQVGF 143
DQ 968 QTG-----QENSDLTTG-YGSTTAGYESSLIAGYGTGTAFKSTLM--AGYG 1013
QY 144 NNATANO 150
DQ 1014 SSQTARE 1020

RESULT 7
ID GP63 LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Button L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729(1988).
RN [2]
RP REVISIONS.
RA Button L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
RA Homans S.W., Bordier C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RT the Leishmania major promastigote surface protease.";
RL J. Biol. Chem. 265:16955-16964(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95406217; PubMed=7675788;
RA Schlagenhauf E., Etges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT Leishmanolysin, the major surface metalloproteinase from Leishmania
RT major.";
RL Proteins 22:58-66(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX MEDLINE=98416698; PubMed=9739094;
RA Schlagenhauf E., Etges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface proteinase
RT Leishmanolysin.";
RL Structure 6:1035-1046(1998).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- PM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, AND C18:0).
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y00647; CAA68673.1; --
 DR PIR; PLO221; PLO221.
 DR PDB; 1LMD; 17-SEP-97.
 DR MEROPS; M08.001; --
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 100 ACTIVATION PEPTIDE.
 FT CHAIN 101 577 LEISHMANOLYSIN.
 FT PROPEP 578 602 REMOVED IN MATURE FORM.
 FT METAL 264 284 ZINC (CATALYTIC).
 FT ACT_SITE 265 285
 FT METAL 269 268
 FT METAL 334 334 ZINC (CATALYTIC).
 FT METAL 125 142 ZINC (CATALYTIC).
 FT DISULFID 191 230
 FT DISULFID 314 386
 FT DISULFID 393 455
 FT DISULFID 406 425
 FT DISULFID 415 489
 FT DISULFID 466 510
 FT DISULFID 515 565
 FT DISULFID 535 558
 FT CARBOHYD 300 300
 FT CARBOHYD 407 407
 FT LIPID 577 577
 FT STRAND 101 102
 FT STRAND 107 108
 FT STRAND 111 114
 FT HELIX 116 119
 FT TURN 121 122
 FT TURN 128 129
 FT STRAND 131 133
 FT STRAND 139 141
 FT HELIX 144 146
 FT HELIX 150 158
 FT TURN 159 159
 FT HELIX 160 169
 FT TURN 170 171
 FT STRAND 172 174
 FT STRAND 177 178
 FT STRAND 180 181
 FT TURN 189 190
 FT HELIX 191 193
 FT HELIX 198 202
 FT TURN 203 203
 FT STRAND 205 206
 FT STRAND 210 215
 FT TURN 221 222
 FT STRAND 226 232
 FT TURN 234 235
 FT STRAND 238 244
 FT HELIX 247 249
 FT HELIX 256 269
 FT TURN 270 271
 FT HELIX 274 279
 FT TURN 280 281
 FT STRAND 283 286
 FT HELIX 289 291
 FT STRAND 296 299
 FT HELIX 302 312

N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 GPI-anchor amidated asparagine.

FT TURN 313 313
 FT TURN 316 316
 FT STRAND 320 322
 FT TURN 328 328
 FT STRAND 335 335
 FT TURN 337 339
 FT TURN 341 342
 FT STRAND 343 343
 FT TURN 344 345
 FT STRAND 353 353
 FT HELIX 356 364
 FT TURN 365 366
 FT STRAND 369 370
 FT HELIX 372 374
 FT TURN 380 383
 FT HELIX 386 390
 FT STRAND 394 395
 FT TURN 396 397
 FT STRAND 398 399
 FT TURN 402 404
 FT STRAND 413 414
 FT TURN 417 418
 FT STRAND 421 425
 FT STRAND 428 429
 FT HELIX 435 437
 FT TURN 443 444
 FT STRAND 445 446
 FT TURN 450 454
 FT STRAND 458 465
 FT TURN 466 467
 FT HELIX 470 472
 FT TURN 475 477
 FT HELIX 478 480
 FT TURN 485 486
 FT STRAND 487 494
 FT STRAND 496 496
 FT STRAND 506 516
 FT TURN 517 520
 FT STRAND 521 525
 FT TURN 527 528
 FT STRAND 533 534
 FT TURN 537 538
 FT STRAND 540 542
 FT HELIX 543 545
 FT TURN 546 546
 FT STRAND 550 550
 FT TURN 552 553
 FT STRAND 555 557
 FT HELIX 561 565
 FT TURN 566 567
 FT HELIX 569 572
 FT TURN 573 573
 SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.8%; Score 91; DB 1; Length 602;
 Best Local Similarity 37.5%; Pred. No. 1.4;
 Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;

QY 63 ARKYDQLVTRVVTHEMAHA-----GQADNSTIELTQNGFRNNATIDOWNAKNSDITV--- 115
 Db 251 ASRYDQLVTRVVTHEMAHALGSGPFFEDARIV-----ANPNVRGKNFDVPEVINS 301
 QY 116 -----GQYG 119
 Db 302 STAVAKAREQYG 313

RESULT 8
 CSGB_SALTI STANDARD; PRT; 151 AA.
 ID_CSGB_SALTI
 AC Q827M3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

```
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR AGFB OR SIM1143.
OS Salmonella typhimurium, and
OC Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]_TaxID=602, 592;
RN SEQUENCE FROM N.A.
RP STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RN [3]
RN J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL627269; CAD08267.1; -
DR EMBL; AE016840; AA069400.1; -
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
Query Match 11.4%; Score 88.5; DB 1; Length 151;
Best Local Similarity 31.0%; Pred. No. 0.48;
Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;
QY 69 LVTRVVTTHMAHAGQAGDNSTIELTQNGFRNNTIDQWNAKNSDIT-VGQYGGNNAALVN 127
Db 48 IIGQVGTDSARVRQEGSKLLSVISQEGGNRAKVDQ--AGNYNFAVIEQTGNANDASIS 105
QY 128 QTASDSSVMVRQVGFGNNAATQY 151
Db 106 QSAYGNSAAIIQKSGNKANITQY 129
RESULT 9
CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

```
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR SIM1143.
OS Salmonella typhimurium, and
OC Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]_TaxID=602, 592;
RN SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=SR-11;
RC MEDLINE=98117058; PubMed=9457880;
RX Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation."
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RX McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RN SEQUENCE FROM N.A.
RP SPECIES=S.enteritidis; STRAIN=27655-3B;
RC MEDLINE=96146512; PubMed=8550497;
RX Collinson S.K., Clouthier S.C., Doran J.L., Banser P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae".
RL J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002301; CAA05316.1; -
DR EMBL; AE008749; AAL20073.1; -
DR EMBL; U43280; AAC43598.1; -
DR PIR; JC6040; JC6040.
DR StyGene; SG10609; csGB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;
Query Match 11.4%; Score 88.5; DB 1; Length 151;
Best Local Similarity 31.0%; Pred. No. 0.48;
Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;
QY 69 LVTRVVTTHMAHAGQAGDNSTIELTQNGFRNNTIDQWNAKNSDIT-VGQYGGNNAALVN 127
Db 48 IIGQVGTDSARVRQEGSKLLSVISQEGGNRAKVDQ--AGNYNFAVIEQTGNANDASIS 105
QY 128 QTASDSSVMVRQVGFGNNAATQY 151
Db 106 QSAYGNSAAIIQKSGNKANITQY 129
```

```

RESULT 10
GP63_LEIDO STANDARD; PRT; 590 AA.
ID P23223;
AC 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
of Leishmania donovani.";
RL Mol. Biochem. Parasitol. 48:173-184 (1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
P1', and basic residues at P2 and P3'. A model nonapeptide is
cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M60048; AAA29244.1; -.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 87 ACTIVATION PEPTIDE.
FT CHAIN 88 565 LEISHMANOLYSIN.
FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 252 252 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 129 BY SIMILARITY.
FT DISULFID 178 217 BY SIMILARITY.
FT DISULFID 301 373 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 565 565 GPI-anchor amidated asparagine (By
similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 63 ARKYDQLVTRVVTHEVAHA 81
Db 238 ASRYDQLVTRVVTHEVAHA 256
| : ||||| ||||| |||||
| : ||||| ||||| |||||

RESULT 11
GP63_LEICH STANDARD; PRT; 599 AA.
ID P15706;
AC 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
Arg-Gly-Asp sequence.";
RL Mol. Biochem. Parasitol. 39:267-274 (1990).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
P1', and basic residues at P2 and P3'. A model nonapeptide is
cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M80672; AAA29238.1; -.
DR EMBL; M28527; AAA29235.1; -.
DR PIR; A44951; A44951.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).

```

Query Match 11.4%; Score 88; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 2.5;

FT METAL 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 BY SIMILARITY.
FT DISULFID 188 BY SIMILARITY.
FT DISULFID 311 BY SIMILARITY.
FT DISULFID 390 BY SIMILARITY.
FT DISULFID 403 BY SIMILARITY.
FT DISULFID 412 BY SIMILARITY.
FT DISULFID 463 BY SIMILARITY.
FT DISULFID 512 BY SIMILARITY.
FT DISULFID 532 BY SIMILARITY.
FT CARBOHYD 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 11.4%; Score 88; DB 1; Length 599;
Best Local Similarity 89.5%; Pred. No. 2.6;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 63 ARKYDQLVTRVVTHEMAHA 81
| : ||||| ||||| |||||
Db 248 ASRYDQLVTRVVTHEMAHA 266
| : ||||| ||||| |||||

RESULT 12
GP63 LEIME STANDARD; PRT; 646 AA.
ID GP63 LEIME STANDARD; PRT; 646 AA.
AC P43150;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin C1 precursor (SC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63-C1
OS Leishmania mexicana.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNYC/BZ/62/M379;
RX MEDLINE=93149206; PubMed=8426614;
RA Medina-Acosta E., Kress R.E., Russell D.G.;
RT "Structurally distinct genes for the surface protease of Leishmania
RT mexicana are developmentally regulated.";
RL Mol. Biochem. Parasitol. 57:31-46(1993).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the
CC amastigote forms.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64394; CAA45733.1; -.
CC PIR; S19916; S19916.
CC HSSP; P08148; 1LML.
CC MEROPS; M08.001; -.
CC GlycoSuiteDB; P43150; -.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase_M8; 1.

DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; Multigene family.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 103 646 LEISHMANOLYSIN C1.
FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 267 267 BY SIMILARITY.
FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 127 144 BY SIMILARITY.
FT DISULFID 193 232 BY SIMILARITY.
FT DISULFID 316 388 BY SIMILARITY.
FT DISULFID 395 458 BY SIMILARITY.
FT DISULFID 408 427 BY SIMILARITY.
FT DISULFID 417 492 BY SIMILARITY.
FT DISULFID 469 513 BY SIMILARITY.
FT DISULFID 518 568 BY SIMILARITY.
FT DISULFID 538 561 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 403 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 646 AA; 69054 MW; FE448DDC78C10B0A CRC64;

Query Match 11.4%; Score 88; DB 1; Length 646;
Best Local Similarity 89.5%; Pred. No. 2.8;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 63 ARKYDQLVTRVVTHEMAHA 81
| : ||||| ||||| |||||
Db 253 ASRYDQLVTRVVTHEMAHA 271
| : ||||| ||||| |||||

RESULT 13
PST3 MYCAV STANDARD; PRT; 369 AA.
ID PST3 MYCAV STANDARD; PRT; 369 AA.
AC QSKK59;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phosphate-binding protein 3 precursor (PBP-3) (PstS-3).
GN PST3.
OS Mycobacterium avium.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=969A45;
RA Carroll J.D., Wallace R.C., Keane J., Arbeit R.D.;
RT "Identification of Mycobacterium avium DNA sequences that encode
RT exported proteins by using phoA gene fusions.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for binding-protein-mediated phosphate
CC transport (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (By similarity).
CC -!- SIMILARITY: Belongs to the pstS family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; AF137360; AAF74819.1; -.
DR HSSP; P06128; 1A54.
DR Inter-Pro; IPR000437; Prok lipoprot_s.
DR Inter-Pro; IPR006059; SBP_bac_1.
DR Pfam; PF01547; SBP_bac_1.
DR PROSITE; PS00013; PROKAP_LIPOPROTEIN; 1.
KW Phosphatase transport; Transport; Membrane; Lipoprotein; Signal;
KW Palmitate.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 369 PHOSPHATE-BINDING PROTEIN 3.
FT LIPID 23 23 N-palmitoyl cysteine (Potential).
FT LIPID 23 23 S-diacetylglucosyl cysteine (Potential).
SQ SEQUENCE 369 AA; 37225 MW; CB0EA0AC10F483EC CRC64;
Query Match 11.2%; Score 86.5; DB 1; Length 369;
Best Local Similarity 27.8%; Pred. No. 2;
Matches 35; Conservative 21; Mismatches 49; Indels 21; Gaps 6;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHNHGGSSGDPSTLSIYQGSANAALALQ 60
Db 1 MKLNFGAVLSVLSAGALVLI-----SGCGSDNNGAGAGAGSSSKVSCGKXKALRASG 54
QY 61 SDARYDQVLTWRV--THEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYG 119
Db 55 STAQA--NATRFVNAFEACPGQ-----TLNYTANG--SGAGISEFNGKQTD-----FG 100
QY 120 GNNAL 125
Db 101 GSDSPL 106
RESULT 14
FME2_ECOLI STANDARD; PRT; 188 AA.
ID FME2_ECOLI STANDARD; PRT; 188 AA.
AC P02972;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE F7-2 fimbrial protein precursor (F7-2 pilin).
GN F7-2 OR PAPA OR C3592.
OS Escherichia coli, and
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;
RN [1]_TaxID=562, 217992;
RP SEQUENCE FROM N.A.
RX MEDLINE=85155489; PubMed=6152241;
RA van Die I., Bergmans H.;
RT "Nucleotide sequence of the gene encoding the F72 fimbrial subunit of
RT a uropathogenic Escherichia coli strain.";
RL Gene 32:183-90(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92040048; PubMed=1682251;
RA Denich K., Blyn L.B., Craiu A., Braaten B.A., Hardy J., Low D.A.,
RA O'Hanley P.D.;
RT "DNA sequences of three papA genes from uropathogenic Escherichia
RT coli strains: evidence of structural and serological conservation.";
RL Infect. Immun. 59:3849-3858(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:Hi / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.I., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley G.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -!- FUNCTION: Fimbriae (also called pili), polar filaments radiating
CC from the surface of the bacterium to a length of 0.5-1.5

CC micrometers and numbering 100-300 per cell, enable bacteria to
CC colonize the epithelium of specific host organs.
CC -!- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12861; AAA23778.1; -.
DR EMBL; M68060; AAA24278.1; -.
DR EMBL; AE016766; AAN82040.1; ALT_INIT.
DR PIR; A03496; YQECF2.
DR Inter-Pro; IPR008966; Adhes_bact.
DR Inter-Pro; IPR000259; Fimbrial.
KW Pfam; PF00419; Fimbrial; 1.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 188 F7-2 FIMBRIAL PROTEIN.
FT DISULFID 43 82 PROBABLE.
SQ SEQUENCE 188 AA; 19184 MW; 0EEF750CFD843157 CRC64;
Query Match 10.9%; Score 84.5; DB 1; Length 188;
Best Local Similarity 25.0%; Pred. No. 1.4; Indels 33; Gaps 8;
Matches 45; Conservative 26; Mismatches 76; Indels 33; Gaps 8;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGHNHGG-----GGNSQDPSTLSIYQYGS 52
Db 2 IKSVIAGAVAMAVSPGAYAPTIPQGGKVTFGTVVDAPCGIDAQADQSIDFGQV-- 59
QY 53 ANAALALQSDA----RKYD-QLVTRVTVTH-ENAHAGQADNSTIELTQNGFRNNATIDQW 106
Db 60 --SKLFLENDGESQPKSFDIKLINCIDITNFKAAGGGGAKTGTSLTSGVPSGQSDML 117
QY 107 ---NAKNSDITVGYQGNNALVNQTSADSSVM-----VROVGFNNATANQY 151
Db 118 QTVGAINTAIVTDPRHGKVKFDGATATGVSILVDGNTIHTFAAVRKGDSGNPVTGAF 177
RESULT 15
ID ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 18-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.


```
CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14992; BAA03636.1; -
DR PIR; JC2143; JC2143.
DR HSRP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 51.
DR PRINTS; PR00327; ICNUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162..993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA22523D333EADD CRC64;

Query Match          10.9%; Score 84.5; DB 1; Length 1034;
Best Local Similarity 26.9%; Pred. No. 9.6;
Matches 45; Conservative 17; Mismatches 66; Indels 39; Gaps 9;

Qy 12 IWVSGSALAGVVPQW--GGGNGHNGGNS-----GPDSTLSIYQYGSNAAL 57
Db 161 IATYGSTLSGTHQSLIAGYGSTETAGDSSTLIAGYGSTGTAGSDSTL-VAGYGSTQTAG 219

Qy 58 ALQSDARKYDQLVTRV----VTHEMAHAGQADNSTIETLQNGFRNNATID----- 104
Db 220 EESSQAGYGSTQTGMKGSDLTAGYGSTGTAGDDSSL-IAGYGSTGTAGDSSSLTAGYGS 278

Qy 105 -QWNAKNSDITVGYGGNNAALVNQTASDSSVMYRVQVFGNNATANQ 150
Db 279 TQTAQKGSDLTAG-YGSTGTA-----GADSSLI---AGYGSTQTAGE 316
```

Search completed: August 2, 2004, 14:49:30
Job time : 6.3 secs

mis Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mac:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 680 | 87.9 | 152 | 2 | O33802 | O33802 salmonella |
| 2 | 582.5 | 75.3 | 150 | 2 | Q7X243 | Q7X243 citrobacter |
| 3 | 552 | 71.3 | 149 | 2 | Q7X240 | Q7X240 citrobacter |
| 4 | 503.5 | 65.1 | 152 | 16 | O8CW63 | O8CW63 escherichia |
| 5 | 430.5 | 55.6 | 150 | 2 | Q7X237 | Q7X237 enterobacte |
| 6 | 301 | 38.9 | 76 | 2 | Q34069 | Q34069 salmonella |
| 7 | 122 | 15.8 | 29 | 2 | Q9S3J5 | Q9S3J5 escherichia |
| 8 | 115 | 14.9 | 130 | 16 | Q89J14 | Q89J14 bradyrhizob |
| 9 | 110 | 14.2 | 139 | 16 | Q8EIH3 | Q8EIH3 shewanella |
| 10 | 106.5 | 13.8 | 502 | 16 | Q8EIH4 | Q8EIH4 shewanella |
| 11 | 105.5 | 13.6 | 145 | 16 | Q8UGN9 | Q8UGN9 agrobacteri |
| 12 | 105.5 | 13.6 | 153 | 16 | Q89J16 | Q89J16 bradyrhizob |
| 13 | 99 | 12.8 | 573 | 10 | Q9SAR2 | Q9SAR2 arabidopsis |
| 14 | 98.5 | 12.7 | 154 | 16 | Q89J15 | Q89J15 bradyrhizob |
| 15 | 97.5 | 12.6 | 1765 | 16 | Q7V8S5 | Q7V8S5 prochloroco |
| 16 | 97 | 12.5 | 157 | 16 | Q88HG0 | Q88HG0 pseudomonas |

| | | | | | | |
|----|------|------|------|----|--------|--------------------|
| 17 | 96.5 | 12.5 | 151 | 16 | Q7UCZ1 | Q7UCZ1 shigella fl |
| 18 | 96.5 | 12.5 | 160 | 16 | Q8CW64 | Q8CW64 escherichia |
| 19 | 96.5 | 12.5 | 160 | 16 | Q83RU7 | Q83RU7 shigella fl |
| 20 | 96 | 12.4 | 2174 | 16 | Q92U08 | Q92U08 rhizobium m |
| 21 | 95 | 12.3 | 1422 | 16 | Q8EFU3 | Q8EFU3 shewanella |
| 22 | 94.5 | 12.2 | 438 | 16 | Q82M56 | Q82M56 streptomyce |
| 23 | 94.5 | 12.2 | 598 | 5 | Q25275 | Q25275 leishmania |
| 24 | 94.5 | 12.2 | 1209 | 16 | Q89CK5 | Q89CK5 bradyrhizob |
| 25 | 93 | 12.0 | 480 | 16 | Q89EV2 | Q89EV2 bradyrhizob |
| 26 | 92.5 | 12.0 | 91 | 2 | Q8S3J8 | Q8S3J8 escherichia |
| 27 | 92.5 | 12.0 | 151 | 2 | Q7X244 | Q7X244 citrobacter |
| 28 | 92 | 11.9 | 644 | 5 | O43994 | O43994 leishmania |
| 29 | 92 | 11.9 | 1408 | 16 | O8E833 | O8E833 shewanella |
| 30 | 91 | 11.8 | 645 | 16 | O7ULC5 | O7ULC5 mycobacteri |
| 31 | 91 | 11.8 | 646 | 16 | O53818 | O53818 mycobacteri |
| 32 | 91 | 11.8 | 1410 | 16 | Q8CM00 | Q8CM00 shewanella |
| 33 | 91 | 11.8 | 2734 | 16 | Q89C73 | Q89C73 bradyrhizob |
| 34 | 90.5 | 11.7 | 329 | 3 | Q8N1Z1 | Q8N1Z1 neurospora |
| 35 | 90 | 11.6 | 152 | 2 | Q7X241 | Q7X241 citrobacter |
| 36 | 90 | 11.6 | 639 | 5 | Q25274 | Q25274 leishmania |
| 37 | 90 | 11.6 | 644 | 5 | O8MNY9 | O8MNY9 leishmania |
| 38 | 90 | 11.6 | 714 | 16 | Q7U5X6 | Q7U5X6 synechococc |
| 39 | 90 | 11.6 | 1460 | 16 | Q8K6B1 | Q8K6B1 streptococc |
| 40 | 90 | 11.6 | 3056 | 16 | Q7USQ0 | Q7USQ0 rhodopirell |
| 41 | 90 | 11.6 | 3501 | 16 | Q8Y106 | Q8Y106 ralstonia s |
| 42 | 90 | 11.6 | 3552 | 16 | Q8XSD6 | Q8XSD6 ralstonia s |
| 43 | 89.5 | 11.6 | 151 | 2 | Q7X238 | Q7X238 enterobacte |
| 44 | 89.5 | 11.6 | 297 | 2 | Q9Z1Y5 | Q9Z1Y5 borrelia bu |
| 45 | 89.5 | 11.6 | 333 | 5 | Q8T986 | Q8T986 drosophila |

ALIGNMENTS

RESULT 1

| | | | |
|--------|---|------|---------|
| O33802 | PRELIMINARY; | PRT; | 152 AA. |
| ID | O33802 | | |
| AC | O33802; | | |
| DT | 01-JAN-1998 (TrEMBLrel. 05, Created) | | |
| DT | 01-JAN-1998 (TrEMBLrel. 05, Last sequence update) | | |
| DT | 01-DEC-2001 (TrEMBLrel. 19, Last annotation update) | | |
| DE | AgfA protein (Fragment). | | |
| GN | AgfA. | | |
| OS | Salmonella typhimurium. | | |
| CC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | |
| CC | Enterobacteriaceae; Salmonella. | | |
| OX | NCBI_TaxID=602; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=98053981; PubMed=9393832; | | |
| RA | Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D., | | |
| RA | Normark S.J., Rhen M.; | | |
| RT | "Expression of thin, aggregative fimbriae promotes interaction of | | |
| RT | Salmonella typhimurium SR-11 with mouse small intestinal epithelial | | |
| RT | cells."; | | |
| RL | Infect. Immun. 65:5320-5325 (1997). | | |
| DR | EMBL; AJ000514; CAA04151.1; -. | | |
| FT | NON_TER | | |
| SQ | SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64; | | |

Query Match 87.9%; Score 680; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 2.3e-48;
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

| | | | | |
|----|-----|---|--------------------------------------|----|
| Qy | 1 | MKLLKVAFAAIVVSGSALA | VVVPQWGGGNGGNSGPDSTLSIYQYGSANAALAQ | 60 |
| Db | 1 | MKLLKVAFAAIVVSGS | VAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALAQ | 60 |
| Qy | 61 | SDARKYDQIVTRVTHEMAHAGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG | 120 | |
| Db | 61 | SDARKSETTITGSGYNGADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG | 120 | |
| Qy | 121 | NNAALVNTASDSSVMVQVGFNNATANQY | 151 | |

```

Db      121 NNAALVNQTASDSVVMVROVGFNNATANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 75.3%; Score 582.5; DB 2; Length 150;
Best Local Similarity 78.8%; Pred. No. 2.3e-40;
Matches 119; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQYGSANAALQ 60
|||||
Db 1 MKLLQVAAIAIVVSGSALAGVVPQWGGG - GGGSSSGPESTLSIYQYGSANAALQ 59
|||||
QY 61 SDARKYDQLVTRVTHEMAHAGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYG 120
|||||
Db 60 SDARKSDTTIHQNGFGNGADVQGGSDNSTIDLTQNGFKNNATIDQWNGKNSDITVSYG 119
|||||

QY 121 NNAALVNQTASDSVVMVROVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSVVLVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 71.3%; Score 552; DB 2; Length 149;
Best Local Similarity 73.5%; Pred. No. 7.2e-38;
Matches 111; Conservative 18; Mismatches 20; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQYGSANAALQ 60
|||||

```

```

Db      1 MKLLKVAAPAAIVVSGSALAGVVPQW - GGNHGGSGNYGPDSSLSIYQYGSNNAALQ 58
|||||
QY 61 SDARKYDQLVTRVTHEMAHAGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYG 120
|||||
Db 59 SDARKSDVTITQHGRNGAVVGGADDSTISLKQTGFQNSATIDQWNAKNSDITVTFGG 118
|||||

QY 121 NNAALVNQTASDSVVMVROVGFNNATANQY 151
|||||
Db 119 RRGALVNQTASDSNVLIQVGFNNATANQH 149
|||||

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT of uropathogenic Escherichia coli.";
RT Extensive mosaic structure revealed by the complete genome sequence
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AEO16759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 65.1%; Score 503.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.1e-34;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW - GGGNGHNGGSGSGDPSTLSIYQYGSANAALQ 59
|||||
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQYGGGGNGHGGGNGNSGPNSELNIYQYGGNSALAQ 60
|||||

QY 60 QSDARKYDQLVTRVTHEMAHAGGADNSTIELTQNGFRNATIDQWNAKNSDITVGYG 119
|||||
Db 61 QADARNSDLTITQHGNGGADVQGGSDSDSIDLTQNGFGNSATIDQWNGKSDTMTVKQFG 120
|||||

QY 120 GNNALVNQTASDSVVMVROVGFNNATANQY 151
|||||
Db 121 GNGAAVQDQTASNSVNVTVQVGFNNATAHQY 152
|||||

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RT Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 55.6%; Score 430.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 6.9e-28;
Matches 89; Conservative 27; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKFIKVAALAIIVSGSANAGWIIQ-GGWGHGHHGGYGGPSTNLINYQGGNSALALQ 59

QY 61 SDARKYDQLVTRVVTHEMAHAGQAGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 TDARNSVLNIGSTGGGNGADVQGGSDSSINLTQNGFNSATLDQWNSKDSVMNVSYGG 119

QY 121 NNAALVNOTASDSSVMVQVQFGNNATANQY 151
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 LNALVDQTASNTVNTQIGFGNHATAHQY 150

RESULT 6
Q954069 PRELIMINARY; PRT; 76 AA.
AC Q954069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGPA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Egelez S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -.
FT NON_TER 76
FT NON_TER 76
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.9%; Score 301; DB 2; Length 76;
Best Local Similarity 80.3%; Pred. No. 1.3e-17;
Matches 61; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 30 GNHNGGSSGPDSTLSIYQYGSANAALALQSDARKYDQLVTRVVTHEMAHAGQAGDNST 89
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGGQAGDNST 60

QY 90 IELTQNGFRNNATIDQ 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q953J5 PRELIMINARY; PRT; 29 AA.
AC Q953J5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Regione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:H80 isolates associated with
RT ISI inserti on in csbs and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.8%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0019;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAATAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q89J14 PRELIMINARY; PRT; 130 AA.
AC Q89J14;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BIL5299 protein.
GN BIL5299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Teurtocka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.9%; Score 115; DB 16; Length 130;
Best Local Similarity 25.9%; Pred. No. 0.046;
Matches 41; Conservative 27; Mismatches 48; Indels 42; Gaps 6;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGG-----GNHNGG-----GNSSGPDSTLSIYQYGS 52
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MRITVLVATAIALSALTVDQAQAGNSASVLQFGTTNSSFSISQGTSTNNATIL---QFGA 57

QY 53 ANAALALQSDARKYDQLVTRVVTHEMAHAGQAGDNSTIELTQNGFRNNATIDQWNAKNSD 112
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 TNTATTLQGS-----LTVNTAVTGGGTAT-----ASNTA 90

QY 113 ITVQYGGNNAALVNOTASDSSVMVQVQFGNNATANQ 150
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 91 LT-GQVGGNSSLIGQIGANNATAGVQGLGLNGSTILQ 127

RESULT 9
Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN SC0866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Felblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.;
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015532; AAN53942.1; -.
DR TIGR; SC0866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.2%; Score 110; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.13;
Matches 30; Conservative 18; Mismatches 34; Indels 24; Gaps 3;

QY 39 SPGSTSLSYQYGSNAALALQSDARKYDQLVTRVVTHEMAHAGQADNSTIETQNGFR 98
DB 41 SGRLNLDLVQQTQNGIFVFS-----GSDNSAY-VTQAGND 77
QY 99 NNATIDQNAKSDITVGYGGNNAALVNOTASDSSVMVROVGFNG 144
DB 78 NISLVQIGT-NNVQLQVGQAKNKASITQIGNDNLVQLNGSGN 122

RESULT 10
Q8EIH4
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Conserved hypothetical protein.
GN SC0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Felblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.;
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015532; AAN53941.1; -.
DR TIGR; SC0865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.8%; Score 106.5; DB 16; Length 502;
Best Local Similarity 23.1%; Pred. No. 1.2;
Matches 40; Conservative 30; Mismatches 72; Indels 31; Gaps 5;

QY 10 AAIWVGSGSALAGVVPWGGG-----GNHNG-----CGNS-----SGPD 42
DB 163 AVERVEGDNDGDIKQYGNNGAGLIADLSANVGNNDVSVIEQIGNNFAGAKGIAGND 222
QY 43 STLSIYQYGSANAALALQSDARKYDQLVTRVVTHEMAHAGQ-CADNSTIETQNGFR-- 99
DB 223 NSVDIVQKGNHGFVYALAGSNDISMEQESGNTAYLSMTTGDNTVDITQDGSNTV 282
QY 100 -NATIDQNAKSDITVGYGGNNAALVNOTASDSSVMVROVGFNGNATANQY 151
DB 283 GDSLIADIQDNDITIKQKGSNGAEFQVWGSNDVDLQKRGDANFATFGAY 335

RESULT 11
Q8U6N9
ID Q8U6N9 PRELIMINARY; PRT; 145 AA.
AC Q8U6N9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Atu4768.
GN ATU4768 OR AGR_L_228.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L.,
RA Chen V., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Iappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Iono C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RA "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AB009405; AAL45562.1; -.
DR EMBL; AB008209; AAK83682.1; -.
DR PIR; AD3143; H98144.
DR PIR; H98144; H98144.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;

Query Match 13.6%; Score 105.5; DB 16; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.32;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY 3 LKVAFAAIVVSGSALAGVVPWGG-----GNHNGGSGSGPDETLSTIYQYSANA 55
DB 1 MIRKSFIALVALVGLSAAAPAMANDVRIEQYGSNAGGAQEGYGNRIPTYQNGYNR 60

```

| | | | |
|---|---|---|-----|
| QY | 56 | ALALQSDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITV | 115 |
| Db | 61 | IVGHQ-----YGR-----HNLAVGQGHGNDYGTSTQNGNRNAGI----- | 96 |
| QY | 116 | GOVGNGNAALVNQTASDSSVMVROVGFGNNTANQ | 150 |
| Db | 97 | GOFGSNHTTILTDQNGNIAAGVQVGRGCSANVSQ | 131 |
| RESULT 12 | | | |
| ID | Q89JI6 | PRELIMINARY; PRT; 153 AA. | |
| AC | Q89JI6; | | |
| DT | 01-JUN-2003 (TrEMBLrel. 24, Created) | | |
| DT | 01-JUN-2003 (TrEMBLrel. 24, Last sequence update) | | |
| DT | 01-JUN-2003 (TrEMBLrel. 24, Last annotation update) | | |
| DE | CsgB protein. | | |
| GN | CSGB OR BLI5297. | | |
| OS | Bradyrhizobium japonicum. | | |
| OC | Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; | | |
| OC | Bradyrhizobiaceae; Bradyrhizobium. | | |
| OX | NCBI_TaxID=375; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=USDA 110; | | |
| RC | MEDLINE=2248498; PubMed=12597275; | | |
| RX | Kaneke T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T., | | |
| RA | Sasanoto S., Watanabe A., Ideawa K., Iriguchi M., Kawashima K., | | |
| RA | Konara M., Macsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M., | | |
| RA | Tabata S.; | | |
| RT | "Complete genomic sequence of nitrogen-fixing symbiotic bacterium | | |
| RT | Bradyrhizobium japonicum USDA110."; | | |
| RL | DNA Res. 9:189-197(2002). | | |
| DR | EMBL; AP005954; BAC50562.1; -. | | |
| KW | Complete proteome. | | |
| SK | SEQUENCE 153 AA; 15991 MW; 4CE7IDEAC375145B CRC64; | | |
| Query Match 13.6%; Score 105.5; DB 16; Length 153; | | | |
| Best Local Similarity 25.7%; Pred. No. 0.34; | | | |
| Matches 39; Conservative 35; Mismatches 55; Indels 23; Gaps 7 | | | |
| QY | 2 | KLLKVA--PAATVVSGLAGVVPQWGGGNGHNGSGSDSTL-SIYQGSANAALAL | 59 |
| Db | 10 | RVLAVALAALGAATQASAGSIQR-----SVTPNVSITETVQFGNDVQPVTI | 58 |
| QY | 60 | QSDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYQY | 119 |
| Db | 59 | EENSRVN---IARV---QIGSGTVDATI--TQNGTRVANVQWGG-TTAAVQSGS | 108 |
| QY | 120 | GNNALVNQTASDSSVMVROVGFGNNTANQY | 151 |
| Db | 109 | LSNTADITQIGNSTNALLIQIGDMNSGAVRQF | 140 |
| RESULT 13 | | | |
| ID | Q9SAF2 | PRELIMINARY; PRT; 573 AA. | |
| AC | Q9SAF2; | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update) | | |
| DT | 01-OCT-2003 (TrEMBLrel. 25, Last annotation update) | | |
| DE | F3F19.21 protein (Hypothetical protein). | | |
| GN | F3F19.21 OR ATIG13190, F3F19.21 OR ATIG13190. | | |
| OS | Arabidopsis thaliana (Mouse-ear cress). | | |
| OC | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | | |
| OC | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; | | |
| OC | eucoids II; Brassicales; Brassicaceae; Arabidopsis. | | |
| OX | NCBI_TaxID=3702; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=cv. Columbia; | | |
| RA | Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lee J.M., Lenz C., | | |
| RA | Liu S., Li J., Kremenetskaia I., Lueros J., Ngan I., Gonzalez A., | | |

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSLA.....DSSVMVRQVFGNNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 776 | 100.0 | 151 | 3 | AAB36351 Agfa::PT3 |
| 2 | 712 | 91.8 | 151 | 3 | AAB36354 Agfa::PT3 |
| 3 | 682 | 87.9 | 151 | 2 | AAR74625 Agfa sequ |
| 4 | 682 | 87.9 | 151 | 3 | AAB36341 Salmonell |
| 5 | 677 | 87.2 | 151 | 2 | AAR23570 Salmonell |
| 6 | 657 | 84.7 | 151 | 3 | AAB36355 Agfa::PT3 |
| 7 | 619 | 79.8 | 151 | 3 | AAB36350 Agfa::PT3 |
| 8 | 609 | 78.5 | 151 | 3 | AAB36352 Agfa::PT3 |
| 9 | 604 | 77.8 | 151 | 3 | AAB36346 Agfa::PT3 |
| 10 | 602 | 77.6 | 151 | 3 | AAB36347 Agfa::PT3 |
| 11 | 601 | 77.4 | 151 | 3 | AAB36353 Agfa::PT3 |
| 12 | 600 | 77.3 | 151 | 3 | AAB36349 Agfa::PT3 |
| 13 | 567 | 73.1 | 151 | 3 | AAB36348 Agfa::PT3 |
| 14 | 521 | 67.1 | 151 | 3 | AAB36343 Escherich |
| 15 | 516 | 66.5 | 151 | 7 | ABR82651 E. coli C |
| 16 | 497 | 64.0 | 120 | 2 | AAR62761 Agfa sequ |
| 17 | 497 | 64.0 | 120 | 2 | AAR23569 Salmonell |
| 18 | 443 | 57.1 | 142 | 2 | AAR52664 Fibronect |
| 19 | 371 | 47.8 | 122 | 2 | AAR52663 FNB curli |
| 20 | 146 | 18.8 | 45 | 3 | AAB36316 Salmonell |
| 21 | 132 | 17.0 | 22 | 3 | AAB36318 Salmonell |
| 22 | 115 | 14.8 | 22 | 3 | AAB36325 Salmonell |
| 23 | 115 | 14.8 | 22 | 3 | AAB36329 Salmonell |
| 24 | 115 | 14.8 | 22 | 3 | AAB36320 Salmonell |
| 25 | 113 | 14.6 | 24 | 7 | ABR82644 E. coli c |

| | | | | | | |
|----|------|------|------|---|----------|---------------------|
| 26 | 111 | 14.3 | 22 | 3 | AAB36322 | Aab36322 Salmonell |
| 27 | 111 | 14.3 | 22 | 3 | AAB36327 | Aab36327 Salmonell |
| 28 | 111 | 14.3 | 22 | 3 | AAB36337 | Aab36337 Salmonell |
| 29 | 109 | 14.0 | 23 | 3 | AAB36340 | Aab36340 Salmonell |
| 30 | 109 | 14.0 | 23 | 3 | AAB36324 | Aab36324 Salmonell |
| 31 | 109 | 14.0 | 23 | 3 | AAB36324 | Aab36324 Salmonell |
| 32 | 102 | 13.1 | 26 | 7 | ABR82649 | ABR82649 E. coli V |
| 33 | 98 | 12.6 | 26 | 7 | ABR82645 | ABR82645 E. coli C |
| 34 | 97.5 | 12.6 | 520 | 6 | AAO16497 | AAO16497 Argiope t |
| 35 | 96 | 12.4 | 19 | 3 | AAB36323 | Aab36323 Salmonell |
| 36 | 96 | 12.4 | 19 | 3 | AAB36336 | Aab36336 Salmonell |
| 37 | 96 | 12.4 | 19 | 3 | AAB36328 | Aab36328 Salmonell |
| 38 | 96 | 12.4 | 2309 | 4 | ABB66232 | ABB66232 Drosophila |
| 39 | 95.5 | 12.3 | 252 | 8 | AD83865 | AD83865 Chemokine |
| 40 | 95 | 12.2 | 597 | 4 | AAU08231 | AAU08231 Polypepti |
| 41 | 94.5 | 12.2 | 738 | 2 | AAW56163 | AAW56163 New DNA s |
| 42 | 92.5 | 11.9 | 151 | 3 | AAB36344 | Aab36344 Escherich |
| 43 | 92.5 | 11.9 | 251 | 5 | ABP45119 | ABP45119 Human BLY |
| 44 | 91 | 11.7 | 1397 | 7 | ADD42761 | ADD42761 Chlamydia |
| 45 | 91 | 11.7 | 1751 | 5 | ABG91039 | ABG91039 Chlamydia |

ALIGNMENTS

RESULT 1

AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
XX
PF 05-APR-2000; 2000WO-CA0000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64627.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAIF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 776; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 9.5e-68;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTTHMAHADQNAKNSDIITVGQYGG 120
 DB 61 SPARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTTHMAHADQNAKNSDIITVGQYGG 120
 QY 121 NNAALVNQTASDSVVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSVVMVRQVGFNNATANQY 151

RESULT 2
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 AC AAB36354;
 XX
 XX 26-FEB-2001 (first entry)
 DT
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 XX WPI: 2000-672631/65.
 DR N-PSDB; AAC64630.
 DR
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal, Agfa
 CC comprising separating an amino acid polymer comprising a recombinant amino
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 91.8%; Score 712; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 1.7e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTTHMAHA-----DOWNAKNSDI 113
 DB 61 SPARKSETTITQSGYNGAD-----YDQLVTRVVTTHMAHAFRNNATIDOWNAKNSDI 113
 QY 114 TVGQYGGNNAALVNQTASDSVVMVRQVGFNNATANQY 151
 DB 114 TVGQYGGNNAALVNQTASDSVVMVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine.
 OS Salmonella.
 XX
 PN WO9425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI; 1994-358275/44.
DR N-PSDB; AAQ87467.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7B; 95pp; English.
XX
CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
CC immunization compositions, respectively, to elicit an immune response to
CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
CC on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 151 AA;

Query Match 87.9%; Score 682; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.4e-58;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYVQGGADNVDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNVDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120

QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 4
AAW23570 standard; protein; 151 AA.
ID AAB36341;
AC AAB36341;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
OS
XX WO200060102-A2.
PN
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
PR
XX (UVVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collinson SK, Kay WW;
PI
XX WPI: 2000-672631/65.
DR
XX N-PSDB; AAC64617.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
PS
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Query Match 87.9%; Score 682; DB 3; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.4e-58;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYVQGGADNVDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNVDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120

QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 5
AAW23570 standard; protein; 151 AA.
ID AAW23570;
AC AAW23570;
XX
XX 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX
XX Salmonella enteritidis 27655-3b agfa.
XX
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
KW Salmonella enteritidis.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 123 /note= "Encoded by GCC"
FT
XX
XX US5635617-A.
PN
XX
XX 03-JUN-1997.
PD
XX 26-APR-1994; 94US-00233788.
PF
XX 26-APR-1993; 93US-00054452.
PR
XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA
XX Collinson SK, Kay WW, Doran JL;
PI


```

PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64626.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 79.8%; Score 619; DB 3; Length 151;
XX Best Local Similarity 74.6%; Pred. No. 2e-52; Indels 44; Gaps 2;
XX Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDSTLSIYQYGSANAALQ 60
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDSTLSIYQYGSANAALQ 60
XX
XX 61 SDARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTHEMAHA----- 103
XX 61 SDARK-----YDQLVTRVVTHEMAHAQQGADNSTIETQNGFR 98
XX
XX 104 -----DOWNAKNSDITVQYGGNNAALVNQTASDSSVVMVROVGFGNNTANQY 151
XX 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVVMVROVGFGNNTANQY 151
XX
XX RESULT 8
XX ID AAB36352
XX XX AAB36352;
XX AC AAB36352;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO2000060102-A2.
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64628.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.5%; Score 609; DB 3; Length 151;
XX Best Local Similarity 73.6%; Pred. No. 1.9e-51;
XX Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
XX
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDSTLSIYQYGSANAALQ 60
XX 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDSTLSIYQYGSANAALQ 60
XX
XX 61 SDARKSETTITQSGYNGADVGQGD-----NYDQLVTRVVT 97
XX 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVTRVT 120
XX
XX 98 HEMAHADQWNAKNSDITVQYGGNNAALVNQTASDSSVVMVROVGFGNNTANQY 151
XX 121 HEMAH-----NOTASDSSVVMVROVGFGNNTANQY 151
XX
XX RESULT 9

```



```

CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 77.4%; Score 602; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 9.3e-51;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
Qy 61 SDARKSETTITQSGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYDQ 120
Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Qy 121 LVTRVVTHEMAHAGVGGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 121 LVTRVVTHEMAHAGVGGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYDQ 120

RESULT 11
AAB36353
ID AAB36353 standard; protein; 151 AA.
AC AAB36353;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX

The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or

```

```

CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;

Query Match 77.4%; Score 601; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.2e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
Qy 61 SDARKSETTITQSGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 61 LVTRVVTHEMAHAGYNGADYDQGVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
XX AAB36349;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX N-PSDB; AAC64625.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:

```

CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 77.3%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 1.5e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADYCGADNYDQLVTRVVTTHMAHADOWNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITQSGYNGADYCGADNSTIETQTQGFNRNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 13
 ID AAB36348
 AC AAB36348; standard; protein; 151 AA.
 DT 26-FEB-2001 (first entry)
 XX AgfA:PT3#3 amino acid sequence SEQ ID NO:16.
 DE Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127889P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.
 CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended are:
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX Sequence 151 AA;

Query Match 73.1%; Score 567; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.4e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 DB 1 MKLLKVAFAAIVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALQ 60
 QY 61 SPARKSETTITQSGYNGADYCGADNYDQLVTRVVTTHMAHADOWNAKNSDITVGOYGG 120
 DB 61 SPARKSETTITQSGYNGADYCGADNSTIETQTQGFNRNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 14
 ID AAB36343
 AC AAB36343; standard; protein; 151 AA.
 DT 26-FEB-2001 (first entry)
 XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 DE Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX Escherichia coli.
 OS WO200060102-A2.
 PN 12-OCT-2000.
 XX


```
PF 05-APR-2000; 2000WO-CA000356.
XX
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVT-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
PI
XX
XX WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 67.1%; Score 521; DB 3; Length 151;
Best Local Similarity 69.5%; Pred. No. 7.5e-43;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAIAAIVFSGSALAGVVPQGGGNGHNGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYGGNGADYVQGDYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120
DB 61 TDARNSDLTITQHGNGGNGADYVQGGSDSDSIDLTQKFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVVRQVGFGNATANQY 151
DB 121 GNGAAVDQTASNSSVNVTVQVGFGNATAHQY 151
RESULT 15
ABR82651
XX ABR82651 standard; protein; 151 AA.
XX
XX ABR82651;
XX
XX 04-DEC-2003 (first entry)
XX
XX E. coli CsgA subunit 15 kDa protein.
XX
```

```
KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX
XX Escherichia coli.
XX
XX WO2003064446-A2.
XX
XX 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX (HANS-) HANSA MEDICAL RES AB.
XX
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-646136/61.
DR N-PSDB; ACP36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
XX Sequence 151 AA;
SQ
Query Match 66.5%; Score 516; DB 7; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.3e-42;
Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVEIAAIVFSGSALAGVVPQGGGNGHNGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYGGNGADYVQGDYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120
DB 61 TDARNSDLTITQHGNGGNGADYVQGGSDSDSIDLTQKFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVVRQVGFGNATANQY 151
DB 121 GNGAAVDQTASNSSVNVTVQVGFGNATAHQY 151
Search completed: August 2, 2004, 14:48:27
Job time : 45.9 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-22
Perfect score: 776
Sequence: 1 MKLLKVAFAAIVWGSALA.....DSSVMVQVGFGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 677 | 87.2 | 151 | 1 | US-08-233-788A-59 |
| 2 | 497 | 64.0 | 120 | 1 | US-08-233-788A-57 |
| 3 | 94.5 | 12.2 | 738 | 3 | US-08-864-038A-3 |
| 4 | 87 | 11.2 | 892 | 4 | US-09-336-447A-5 |
| 5 | 86.5 | 11.1 | 238 | 4 | US-09-495-880A-42 |
| 6 | 86.5 | 11.1 | 585 | 4 | US-09-620-412C-337 |
| 7 | 86.5 | 11.1 | 585 | 4 | US-09-598-419-337 |
| 8 | 86.5 | 11.1 | 1752 | 4 | US-09-556-877-180 |
| 9 | 86.5 | 11.1 | 1752 | 4 | US-09-620-412C-180 |
| 10 | 86.5 | 11.1 | 1752 | 4 | US-09-598-419-180 |
| 11 | 85.5 | 11.0 | 589 | 4 | US-09-498-039A-7849 |
| 12 | 82 | 10.6 | 975 | 4 | US-09-328-352-4764 |
| 13 | 80 | 10.3 | 461 | 4 | US-09-252-991A-24717 |
| 14 | 80 | 10.3 | 812 | 4 | US-09-252-991A-30710 |
| 15 | 80 | 10.3 | 1034 | 4 | US-09-252-991A-26658 |
| 16 | 79 | 10.2 | 943 | 4 | US-09-056-556-204 |
| 17 | 79 | 10.2 | 943 | 4 | US-09-072-596-199 |
| 18 | 79 | 10.2 | 943 | 4 | US-09-477-135A-131 |
| 19 | 79 | 10.2 | 943 | 4 | US-09-072-967-204 |
| 20 | 79 | 10.2 | 2315 | 4 | US-09-543-681A-5434 |
| 21 | 78.5 | 10.1 | 209 | 4 | US-09-125-619-42 |
| 22 | 78.5 | 10.1 | 278 | 3 | US-09-260-283-2 |
| 23 | 78.5 | 10.1 | 309 | 4 | US-09-252-991A-22266 |
| 24 | 78.5 | 10.1 | 745 | 4 | US-09-336-115C-6 |
| 25 | 78 | 10.1 | 392 | 2 | US-08-387-942C-23 |
| 26 | 78 | 10.1 | 553 | 2 | US-08-387-942C-2 |
| 27 | 78 | 10.1 | 645 | 4 | US-09-919-172-41 |

| | | | | | | |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 78 | 10.1 | 878 | 4 | US-09-540-236-3401 | Sequence 3401, Ap |
| 29 | 78 | 10.1 | 1216 | 4 | US-09-134-000C-5130 | Sequence 5130, Ap |
| 30 | 77.5 | 10.0 | 212 | 4 | US-09-125-619-32 | Sequence 32, Appl |
| 31 | 77.5 | 10.0 | 437 | 2 | US-08-737-716-2 | Sequence 2, Appl |
| 32 | 77.5 | 10.0 | 673 | 3 | US-09-196-387-8 | Sequence 8, Appl |
| 33 | 77.5 | 10.0 | 673 | 4 | US-09-841-835-8 | Sequence 8, Appl |
| 34 | 77.5 | 10.0 | 702 | 4 | US-09-252-991A-22119 | Sequence 22119, A |
| 35 | 77.5 | 10.0 | 941 | 4 | US-09-336-447A-9 | Sequence 9, Appl |
| 36 | 77.5 | 10.0 | 949 | 3 | US-09-196-387-10 | Sequence 10, Appl |
| 37 | 77.5 | 10.0 | 949 | 4 | US-09-841-835-10 | Sequence 2, Appl |
| 38 | 77.5 | 10.0 | 1327 | 3 | US-09-196-387-2 | Sequence 2, Appl |
| 39 | 77.5 | 10.0 | 1327 | 4 | US-09-841-835-2 | Sequence 8, Appl |
| 40 | 77.5 | 10.0 | 1327 | 4 | US-09-972-115A-8 | Sequence 2, Appl |
| 41 | 77.5 | 10.0 | 1690 | 4 | US-09-595-684B-39 | Sequence 39, Appl |
| 42 | 77 | 9.9 | 266 | 4 | US-09-495-880A-26 | Sequence 26, Appl |
| 43 | 77 | 9.9 | 339 | 4 | US-09-252-991A-32096 | Sequence 32096, A |
| 44 | 77 | 9.9 | 873 | 4 | US-09-336-447A-13 | Sequence 13, Appl |
| 45 | 77 | 9.9 | 1139 | 1 | US-08-537-210A-4 | Sequence 4, Appl |

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 87.2%; Score 677; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 3.3e-59;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVWGSALAGVVPWGGGNGNGSGPDTLSIYQYGSANAALAQ 60
|||||

| | | | |
|----|-----|---|-----|
| Db | 1 | MKLLKVAAPAAITVSSGSALAGVVPQWGGGNGHNGSGSPDSTLSITYYGSANAALAIQ | 60 |
| Qy | 61 | SPARKSETTITQSGYNGGADVYQGGADNDVLVTRVVTHEMAHADQWNKNSDITVQYQGG | 120 |
| Db | 61 | SPARKSETTITQSGYNGGADVYQGGADNSTLTQTQGFNNATIDQWNKNSDITVQYQGG | 120 |
| Qy | 121 | NNAALVNQTASDSSVMVRQVGFEGNNATANQY | 151 |
| Db | 121 | NNPALVNQTASDSSVMVRQVGFEGNNATANQY | 151 |

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue

```

Query Match      64.0%; Score 497; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 1.2e-41;
Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

22 VVPQWGGGHNHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYGCADV 81
|||||
1 VVPQWGGGHNHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYGCADV 60
|||||

82 GQGADNDQLVTRVVRTHMAHADOWNAKNSDITVQYCGNNAALVQNTASD 133
|||||
61 GQGADNSTIITQGRFNNAITDWNKNSDITVQYCGNNAALVQNTASD 112
|||||

```

RESULT 3
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:

APPLICANT: Kunio NAKASHIMA et al.
 TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
 TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
 TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
 TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
 TITLE OF INVENTION: TO SAID POLYPEPTIDE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: 812-5 Hirano
 STREET: Isshinden
 CITY: Tsu-city
 STATE: Mie-prefecture
 COUNTRY: JAPAN
 ZIP: 514-01
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Microsoft Windows 95
 SOFTWARE: Word Perfect 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/864,038A
 FILING DATE: May 28, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-184459
 FILING DATE: 15-July-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: C. Bruce Hamburg
 REGISTRATION NUMBER: 22,389
 REFERENCE/DOCKET NUMBER: F-5610
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)986-2340
 TELEFAX: (212)953-7733
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 738
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Pinctada fucata
 CELL TYPE: mantle epithelial cell
 FEATURE:
 NAME/KEY: peptide
 LOCATION: from 1 to 738
 IDENTIFICATION METHOD: E (by experiment)
 US-08-864-038A-3

RESULT 4
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL

```
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Query Match      11.2%; Score 87; DB 4; Length 892;
Best Local Similarity 26.8%; Pred. No. 3.4;
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNHN-----GGNNGSGPDTLSIYQYGSANAALALQSDARKSETTITQSGY--- 75
Db 89 GGGDYNEAKGNYSIVGGSSNTAKGSKSTIGGGDTN-----DANGTYSITGGYYSRA 141
QY 76 -GNGADVQGGADNYDLVTRVVTHEMAHADQNAKSDITVQYG---GNNAAALV-----N 127
Db 142 IGSSTTIGGG--YINQATGEKSTVAGGRNNQATGNNSTVAGGSYNOATGNNSTVAGGSHN 199
QY 128 QTASDSSVMVVRQVGFGNNTAN 149
Db 200 QATGEGSF---AAGVENKANAN 218

RESULT 5
US-09-495-880A-42
; Sequence 42, Application US/09495880A
; Patent No. 6667150
; GENERAL INFORMATION:
; APPLICANT: RUDERT, FRITZ
; APPLICANT: GE, LIMING
; APPLICANT: ILAG, VIC
; TITLE OF INVENTION: NOVEL METHOD AND PHASE FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING MEMBERS OF A MULTIMERIC
; TITLE OF INVENTION: (POLY) PEPTIDE COMPLEX
; FILE REFERENCE: MORPHO/9
; CURRENT APPLICATION NUMBER: US/09/495,880A
; CURRENT FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: PCT/EP98/04836
; PRIOR FILING DATE: 1998-08-03
; PRIOR APPLICATION NUMBER: EP 97 11 3319.4
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 42
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: ompA-FLAG-peptide3-
; OTHER INFORMATION: gene I11s encoded by phage vector fpep3_1B-ir3seq (circular)
US-09-495-880A-42

Query Match      11.1%; Score 86.5; DB 4; Length 238;
Best Local Similarity 24.3%; Pred. No. 0.69;
Matches 43; Conservative 24; Mismatches 63; Indels 47; Gaps 8;

QY 5 KVAAPAAIVVSGSALAG-----VVPGWGGGNH--NGGNGSSGPDSTLSI 47
Db 3 KTAIAIALAGFATVQAQYKVDVCIVYHAHLVAKCGGSGSEFNAGGSGG----- 55
QY 48 YQYGSANAALQSDARKSETTITQSGYNGADYVQGGADNYDLVTRVVTHEMAHADQNA 107
Db 56 ---GSGGSEGGEGGEGGEGGEGG--GSGGSGSGGDFDYKMKANANKAMTENADE-N 110
QY 108 AKNSDI-----TVQYQG-----GNNAAALVNQTA-----SDSSVMVRQVGFGNNA 146
```

```
Db 111 ALQSDAKGLSDVATDYGAAIDGFIGDYVGLANGATGDFAGNSQMAQVQGDGNS 167

RESULT 6
US-09-620-412C-337
; Sequence 337, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
; APPLICANT: Steven P. Fling
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C7
; CURRENT APPLICATION NUMBER: US/09/620,412C
; CURRENT FILING DATE: 2000-07-20
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-620-412C-337

Query Match      11.1%; Score 86.5; DB 4; Length 585;
Best Local Similarity 31.6%; Pred. No. 2.2;
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAAPAAIVVSGSALAGVVPQWGGGN--HNGGNGSSGPDSD---TLSIYQYGSANAAL 57
Db 206 LQAQASAG---NADAWASSSPQSGGATTVSDSGSSGSDTSETVPTAKGG----- 257
QY 58 ALQSDARKSETTIT---QSGYNGADYVQGGADNYDLVTRVVTHEMAHADQNAKNSDIT 114
Db 258 GLYTDKNLSITNITGIIIEIANNAKATDVGGGA-----YVKGTLTCENSHRLQFLKNSDDQ 312
QY 115 VQY-GGNNAAALVNQT 129
Db 313 GGGIYGEONITLSNLT 328

RESULT 7
US-09-598-419-337
; Sequence 337, Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Scholler, John
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 337
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-09-598-419-337

Query Match      11.1%; Score 86.5; DB 4; Length 585;
Best Local Similarity 31.6%; Pred. No. 2.2;
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAAPAAIVVSGSALAGVVPQWGGGN--HNGGNGSSGPDSD---TLSIYQYGSANAAL 57
Db 206 LQAQASAG---NADAWASSSPQSGGATTVSDSGSSGSDTSETVPTAKGG----- 257
QY 58 ALQSDARKSETTIT---QSGYNGADYVQGGADNYDLVTRVVTHEMAHADQNAKNSDIT 114
Db 258 GLYTDKNLSITNITGIIIEIANNAKATDVGGGA-----YVKGTLTCENSHRLQFLKNSDDQ 312
QY 115 VQY-GGNNAAALVNQT 129
```

Db 440 GLYTDKMLSTNTGTGIIIEIANNKATDVGGGA-----YVKGTLTCTENSHRLQFLKNSSDKQ 494

Qy 115 VGO-YGGNNAALVNQT 129
 : : : : :
 Db 495 GGGIYGEDNITLSNLT 510

RESULT 10
 US-09-598-419-180
 ; Sequence 180, Application US/09598419
 ; Patent No. 6565856
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Scholler, John
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
 ; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
 ; FILE REFERENCE: 210421.469C6
 ; CURRENT APPLICATION NUMBER: US/09/598,419
 ; CURRENT FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 357
 ; SOFTWARE: FastSeq for Windows Version 3.0/4.0
 ; SEQ ID NO 180
 ; LENGTH: 1752
 ; TYPE: PRT
 ; ORGANISM: Chlamydia
 US-09-598-419-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;
 Best Local Similarity 31.8%; Pred. No. 9;
 Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

Qy 4 LKVAFAAIVVSGSALAGVVPQWGGGN--HNGGNGSSGPDG----TLSIYQYGSANAAL 57
 : : : : :
 Db 388 LKAQASAG--NADAWASSPQSGSGATTVDSGDSGSDTSETVPTAKGG----- 439

Qy 58 ALQSDARKSETTIT---QSGYNGADVQGGADNYDQLVTRVVTHEMAHADQWNAKNSDIT 114
 : : : : :
 Db 440 GLYTDKMLSTNTGTGIIIEIANNKATDVGGGA-----YVKGTLTCTENSHRLQFLKNSSDKQ 494

Qy 115 VGO-YGGNNAALVNQT 129
 : : : : :
 Db 495 GGGIYGEDNITLSNLT 510

RESULT 11
 US-09-489-039A-7849
 ; Sequence 7849, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 7849
 ; LENGTH: 589
 ; TYPE: PRT
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-7849

Query Match 11.0%; Score 85.5; DB 4; Length 589;
 Best Local Similarity 23.2%; Pred. No. 2.8;
 Matches 32; Conservative 17; Mismatches 68; Indels 21; Gaps 3;

Qy 29 GGNHNGGNGSSGPDSTLSIYQYS-----ANAAALQSDARKSETTITQSGYNGAD 80
 : : : : :
 Db 259 GGLDRNGANAGQDTFGIYAFDTLTLTRIEINGRLDNYTKYDSATACGSGRGAI 318
 : : : : :
 Qy 81 VQGGADNYDQLVTRVVTHEMAHADQWNA-----KNSDITVQYGGNNAALVNQTASDS 133

Db 319 ACPRGQSTGSPVTTVDTAQSNLVNWKAGALYRLTEQNVYV-----NYAISQPPGGS 372
QY 134 SVMVRQVGFGNATANQY 151
Db 373 SPALAAAGSGNSANRTRDF 390

RESULT 12
US-09-328-352-4764
; Sequence 4764, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 10.6%; Score 82; DB 4; Length 975;
Best Local Similarity 24.1%; Pred. No. 12;
Matches 38; Conservative 15; Mismatches 47; Indels 58; Gaps 7;
QY 15 GSGALAGVVPWGGGNGHNGG-GNSSGPDSTLSIYQYG-----SANA 55
Db 300 AGNGIA-----SGNGEYHNGGNGGDDVDITAPITGLNLSGNSFTLNGSSSSSVNT 353
QY 56 ALALOSDARKSETTI-----TQSGYG-----NGADVGQAGADNDYDQLVTRVWTHM 100
Db 354 APTTTSNTVNDNTIDNGSGGTGSGNGSGDGLLNGAASNGEYHNG----- 402
QY 101 AHADOWNAKNSDIT-----VGQYGGNNAALVNQTSADS 133
Db 403 --IGNGNGDDVDITSPITGIFNFGNSFSLGNSSSSS 438

RESULT 13
US-09-252-991A-24717
; Sequence 24717, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24717
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24717

Query Match 10.3%; Score 80; DB 4; Length 461;
Best Local Similarity 25.6%; Pred. No. 7;
Matches 43; Conservative 20; Mismatches 67; Indels 38; Gaps 8;
QY 6 VAAPAAIVSG-SALAGVVPWGGGNGHNGGNSGPDSTLSIYQVGSANA---ALALQS 61
Db 300 VAALPEVARESGAPSGTAPAGGAA---GGKSPAGLRLARRKSPSSATPPAATTYF 356

QY 62 DARKSETTITQSG-YCNGA-----DVGGADNYDQVLT-----RVVTHEMAHADOWN- 107
Db 357 AARAPPAITEPRSGTGAPDPRRTAGTCEQGLVVDFAHPAYRLVLTGQIEGDSWNG 416
QY 108 -----AKNSDITVQYGGNNAALVNQTSADSSVMVRQVGFGNNA 146
Db 417 DVFLRIDMAAELOQDFMGADGNQAVASQVISET-----VGNHA 456

RESULT 14
US-09-252-991A-30710
; Sequence 30710, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30710
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30710

Query Match 10.3%; Score 80; DB 4; Length 812;
Best Local Similarity 26.7%; Pred. No. 15;
Matches 31; Conservative 10; Mismatches 39; Indels 36; Gaps 5;
QY 51 GSAALALALQSDARKSE-----TTTQSGYNGADVGGADNYDQLVTRVTHEMAHAD 104
Db 550 GLNIGVTRDSRRYSERVIVSRSTPSQGLGNLGYGGGASRYQQ-----AD 597
QY 105 -QWAKNSDITVQYG--GN-----NAALVNQTSADSSVMVRQVGF 142
Db 598 LTWRMQNVLOGGLYGETGNYTRWADLSGLVMDNAVFASNRINDAFVLVSTKGY 653

RESULT 15
US-09-252-991A-26658
; Sequence 26658, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26658
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26658

Query Match 10.3%; Score 80; DB 4; Length 1034;
Best Local Similarity 26.7%; Pred. No. 20;
Matches 47; Conservative 19; Mismatches 62; Indels 48; Gaps 10;
QY 14 VSGALAGVVPWGGGNGHNG-----GGNSGPDSTLSIYQVGS-ANAALALQSDAR-- 64
Db 748 VDSASASQV-EAGAGNTGLVGLSSGGEIFRQASGVSKGLATGGLIGKAEGNGM 806

```
Qy 65 ----KSETTITQSGYNGADVGGADNYDQLV-----TRY----- 95
Db 807 LGNLKASGSVTDQG---GADLGLVGNNSQSAIETAEATGKVS GGSNSRVGGGLIGHNLGG 863
Qy 96 -VTHEMAHADQWNAKNSDITVGYGGNNAALVNOTASD--SSVMVRQVG--FGNNA 146
Db 864 SVAHAI SRGDVSGGFNS-LVGGLVGHNGGELVNVDSGRVSAASASVGGLVGSNA 918
```

Search completed: August 2, 2004, 14:58:35
Job time : 13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------------|--------------------|
| 1 | 518 | 66.8 | 151 | 12 US-09-741-873B-4 | Sequence 4, Appli |
| 2 | 518 | 66.8 | 151 | 12 US-09-741-873B-4 | Sequence 4, Appli |
| 3 | 440 | 56.7 | 131 | 12 US-09-741-873B-2 | Sequence 2, Appli |
| 4 | 440 | 56.7 | 131 | 12 US-09-741-873B-2 | Sequence 2, Appli |
| 5 | 95 | 12.2 | 445 | 15 US-10-369-493-20638 | Sequence 20638, A |
| 6 | 95 | 12.2 | 597 | 9 US-09-793-306-146 | Sequence 146, App |
| 7 | 94 | 12.1 | 271 | 16 US-10-437-963-147343 | Sequence 147343, |
| 8 | 93 | 12.0 | 369 | 12 US-10-425-114-56041 | Sequence 56041, A |
| 9 | 93 | 12.0 | 486 | 12 US-10-424-599-275468 | Sequence 275468, |
| 10 | 93 | 12.0 | 507 | 12 US-10-425-114-57763 | Sequence 57763, A |
| 11 | 92.5 | 11.9 | 251 | 10 US-09-880-748-1130 | Sequence 1130, Ap |
| 12 | 92.5 | 11.9 | 251 | 12 US-10-293-418-1130 | Sequence 1130, Ap |
| 13 | 91 | 11.7 | 1751 | 9 US-09-841-132-445 | Sequence 445, App |
| 14 | 91 | 11.7 | 1751 | 9 US-09-841-132-594 | Sequence 594, App |
| 15 | 91 | 11.7 | 1751 | 16 US-10-467-534-45 | Sequence 45, Appli |

89:5 11.5 251 10 US-09-880-748-1122 Sequence 1122, Ap
89:5 11.5 251 12 US-10-293-418-1122 Sequence 1122, Ap
89 11.5 193 16 US-10-437-963-148500 Sequence 148500,
18 11.5 145 16 US-10-437-963-147748 Sequence 147748,
19 11.4 145 16 US-10-425-114-67750 Sequence 67750, A
20 88:5 11.4 197 12 US-10-437-963-158876 Sequence 158876,
21 88 11.3 503 16 US-10-437-963-162284 Sequence 162284,
22 87:5 11.3 154 16 US-10-437-963-162284 Sequence 162284,
23 87:5 11.3 204 12 US-10-424-599-203972 Sequence 203972,
24 87:5 11.3 486 15 US-10-369-493-20619 Sequence 20619, A
25 87:5 11.3 689 16 US-10-437-963-188971 Sequence 188971,
26 87 11.2 892 10 US-09-952-267-5 Sequence 5, Appli
27 86:5 11.1 191 16 US-10-437-963-105413 Sequence 105413,
28 86:5 11.1 238 12 US-10-634-862-42 Sequence 42, Appli
29 86:5 11.1 585 9 US-09-841-132-337 Sequence 337, App
30 86:5 11.1 1752 9 US-09-841-132-180 Sequence 180, App
31 86 11.1 448 16 US-10-437-963-140685 Sequence 140685,
32 86 11.1 735 12 US-10-425-114-63965 Sequence 63965, A
33 86 11.1 6310 12 US-10-282-122A-67793 Sequence 67793, A
34 85:5 11.0 270 16 US-10-437-963-122263 Sequence 122263,
35 85:5 11.0 580 12 US-10-647-057-4 Sequence 4, Appli
36 85 11.0 354 10 US-09-820-843A-21 Sequence 21, Appli
37 85 11.0 400 16 US-10-437-963-186417 Sequence 186417,
38 84:5 10.9 189 16 US-10-437-963-170736 Sequence 170736,
39 84:5 10.9 1276 16 US-10-437-963-168952 Sequence 168952,
40 84 10.8 242 12 US-10-425-114-61520 Sequence 61520, A
41 84 10.8 253 10 US-09-880-748-2098 Sequence 2098, Ap
42 84 10.8 253 12 US-10-293-418-2098 Sequence 2098, Ap
43 84 10.8 255 10 US-09-880-748-1153 Sequence 1153, Ap
44 84 10.8 255 12 US-10-293-418-1153 Sequence 1153, Ap
45 84 10.8 1448 16 US-10-408-765A-998 Sequence 998, App

ALIGNMENTS

RESULT 1
US-09-741-873B-4
; Sequence 4, Application US/09741.873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741.873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.8%; Score 518; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-44;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPSTLSIYQYGSANAALQ 60

```
Db      1  MLLKVAATAAIVFSSAVAGVVPQGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY      61  SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db      61  TDARNSDLTITQGGGNGADVGQSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNOTASDSSVMVROVQFGNNATANQY 151
Db      121  GNGAAVDQIASNSSVNVTVQFGNNATAHQY 151
```

RESULT 2

```
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4
```

```
Query Match      66.8%; Score 518; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-44;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1  MLLKVAATAAIVFSSALAGVVPQGGGNGHGGGNSGPDSTLSIYQYGSANAALALQ 60
Db      1  MLLKVAATAAIVFSSAVAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY      61  SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db      61  TDARNSDLTITQGGGNGADVGQSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNOTASDSSVMVROVQFGNNATANQY 151
Db      121  GNGAAVDQIASNSSVNVTVQFGNNATAHQY 151
```

RESULT 3

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
```

```
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
```

```
Query Match      56.7%; Score 440; DB 12; Length 131;
Best Local Similarity 65.8%; Pred. No. 8.3e-37;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

QY      21  GVVPQGGGNGHGGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db      1  GVVPYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLTITQGGGNGAD 60
QY      81  VGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNOTASDSSVMVROV 140
Db      61  VGQSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGGNGAAVDQIASNSSVNVTVQV 120
QY      141  GFGNNATANQY 151
Db      121  GFGNNATAHQY 131
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match      56.7%; Score 440; DB 12; Length 131;
```

[illegible]

```

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodopseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

```

| | | | | |
|-----------------------|-------|--|---------------|------------|
| Query Match | 12.2% | Score 95 | DB 15 | Length 445 |
| Best Local Similarity | 24.6% | Pred. No. 0.41 | | |
| Matches | 35 | Conservative 22 | Mismatches 59 | Indels 26 |
| Gaps | 5 | | | |
| Qy | 30 | GNHGGGNSS-----GPDSTLSIYQGSANAA----- | LALQSDARKSET | 68 |
| Db | 89 | GKSGAGNSAALFQGTGSDVELOQTGTSNAGVPSGWNWTDPGVFNKLTQSSSNGSKV | 148 | |
| Qy | 69 | TITQSGYNGADVGQGANLYQLVTRVVTHEMAHADQW-NAKNSDITVQQ---YGGN-NA | 123 | |
| Db | 149 | SVIQDGKNNVFIKQGTNGNSTSVNQIGEWGMAVVRQIGAAETDASTGNALPTGNNYV | 208 | |
| Qy | 124 | ALVNTQASDSSVNVVRQVFGNN | 145 | |
| Db | 209 | ASITONSAGLNVAVAVQGGNS | 230 | |

RESULT 6
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Overdale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

```

; TITLE OF INVENTION: Of Tuberculosis
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mtTC#3-His
US-09-793-306-146

```

```

Query Match      12.2%; Score 95; DB 9; Length 597;
Best Local Similarity 27.4%; Pred. No. 0.59;
Matches 34; Conservative 14; Mismatches 50; Indels 26; Gaps 4
QY    26 WGGGNNHNGGNSGSPDSTLSIYYGSANAALAQSDARKSETTITQSgyGNAGdyQGQA 85
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     358 FNSGNNNIIGFNSG-NNNVGFNFGSNN-----FFGNGADINTGF 398
QY    86 DNYDLVTRVYTHEMAHADOWNAKNSDITYGYQyGNNNAALVNQTASDSSVMVRQyVGfGNN .45
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     399 GNAGDTNTGFCGNAGFNNGIGNAGNEDMGyGVNGGSFNVyGVN--AGNQS----VGfGNA 451
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
QY    146 ATAN 149
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     452 GTLN 455

```

```

RESULT 7
US-10-437-963-147343
; Sequence 147343, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221) B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147343
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; US-10-437-963-147343
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47881C.1.pep
; US-10-437-963-147343

```

| | | | | |
|-----------------------|--------|------------------|----------------|-------------------|
| Query Match | 12.1% | Score 94; | DB 16; | Length 271; |
| Best Local Similarity | 26.2%; | Pred.No. 0.28; | | |
| Matches | 37; | Conservative 19; | Mismatches 67; | Indels 18; Gaps 6 |

| | | | |
|----|----|---|-----|
| QY | 14 | VSGSALGVVQWGGG-GNNHGGNSGGPSTLSIYQYGSANAALALQSDARKSETTTQ | 72 |
| | | | |
| Db | 7 | VYGRAAAAAYPEVSGDGSGSGGGGGGG------GGDGSVAAVNFAGSGGDCRSS | 58 |
| | | | |
| QY | 73 | SGYGNADVGGQADNYDGLVTVVTHFMADQWNAKNSDITVQYCGGNAALVQNTASD | 132 |
| | | | |
| Db | 59 | GGEQGGSSGGGLGRQRYNRSLSRTEORLVDH-VFKNSDVQCDVFSG-----VGPATS | 112 |
| | | | |

```

QY 133 SVMVROVGFGR--NATANQY 151
Db 113 AARKVRIV-YANDLNPTAVEY 132

RESULT 8
US-10-425-114-56041
; Sequence 56041, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56041
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701205720_FLI.pep
US-10-425-114-56041

Query Match 12.0%; Score 93; DB 12; Length 369;
Best Local Similarity 26.3%; Pred. No. 0.51;
Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETITQS 73
Db 13 VRSSASKG-KPPLSAGNNNGGNDWDDPD-----GFGSARGG-----ADLRNQSTGDVR 62

QY 74 GYGNAGADVGGADNYDQLVTRVVTTHMAHADQWNAKNSDITVGYGNNALVNQTASDS 133
Db 63 GFGGGNVRSKSTQD-MYTR-----AELEASAANKEDFFAKRAENESRPEGLPPSQG 115

QY 134 SVMVROVGFGRNATANQ 150
Db 116 G---KYVFGSGGAPNQ 129

RESULT 9
US-10-424-599-275468
; Sequence 275468, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 275468
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90769C.1.pep
US-10-424-599-275468

Query Match 12.0%; Score 93; DB 12; Length 486;
Best Local Similarity 26.3%; Pred. No. 0.72;

```

```

Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETITQS 73
Db 130 VRSSASKG-KPPLSAGNNNGGNDWDDPD-----GFGSARGG-----ADLRNQSTGDVR 179

QY 74 GYGNAGADVGGADNYDQLVTRVVTTHMAHADQWNAKNSDITVGYGNNALVNQTASDS 133
Db 180 GFGGGNVRSKSTQD-MYTR-----AELEASAANKEDFFAKRAENESRPEGLPPSQG 232

QY 134 SVMVROVGFGRNATANQ 150
Db 233 G---KYVFGSGGAPNQ 246

RESULT 10
US-10-425-114-57763
; Sequence 57763, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 57763
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY049H04_FLI.pep
US-10-425-114-57763

Query Match 12.0%; Score 93; DB 12; Length 507;
Best Local Similarity 26.3%; Pred. No. 0.76;
Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETITQS 73
Db 151 VRSSASKG-KPPLSAGNNNGGNDWDDPD-----GFGSARGG-----ADLRNQSTGDVR 200

QY 74 GYGNAGADVGGADNYDQLVTRVVTTHMAHADQWNAKNSDITVGYGNNALVNQTASDS 133
Db 201 GFGGGNVRSKSTQD-MYTR-----AELEASAANKEDFFAKRAENESRPEGLPPSQG 253

QY 134 SVMVROVGFGRNATANQ 150
Db 254 G---KYVFGSGGAPNQ 267

RESULT 11
US-09-880-748-1130
; Sequence 1130, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16

```

```
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1130
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1130

Query Match      11.9%; Score 92.5; DB 10; Length 251;
Best Local Similarity 27.7%; Pred. No. 0.36;
Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;

QY 18 ALAGVVPWG-----GGGNHGGGSSGPDSTLSIYQYGSANAALALQSDARKSE- 67
DB 106 ATTGALDMWGKTLVTVSSGGGGGGGG-----GSAQAVLTQPSASGTPG 155
QY 68 TTITOSGYNGADVGQADN-YDQLVTRVVTHEMAHADQWNAKNSDITVGOYGGNNAAL 125
DB 156 QRVWMSGSSSNGSNTVNWYQQLPGAAPKLLIYRSQRRSGVDPDRFSGSKSGTSASL 214

RESULT 12
US-10-293-418-1130
; Sequence 1130, Application US/10293418
; Publication No. US20030223996A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523P2
; CURRENT APPLICATION NUMBER: US/10/293,418
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/331,469
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 60/340,817
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 09/880,748
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 3247
; SEQ ID NO 1130
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-418-1130

Query Match      11.9%; Score 92.5; DB 12; Length 251;
Best Local Similarity 27.7%; Pred. No. 0.36;
Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;

QY 18 ALAGVVPWG-----GGGNHGGGSSGPDSTLSIYQYGSANAALALQSDARKSE- 67
DB 106 ATTGALDMWGKTLVTVSSGGGGGGGG-----GSAQAVLTQPSASGTPG 155
QY 68 TTITOSGYNGADVGQADN-YDQLVTRVVTHEMAHADQWNAKNSDITVGOYGGNNAAL 125
DB 156 QRVWMSGSSSNGSNTVNWYQQLPGAAPKLLIYRSQRRSGVDPDRFSGSKSGTSASL 214

RESULT 13
US-09-841-132-445
```

```
; Sequence 445, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 445
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis serovar D
US-09-841-132-445

Query Match      11.7%; Score 91; DB 9; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SGSALAGVVPWGCGGN--HNGGNSGSGPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
DB 394 NADAWASSPQSGGATTVNSGDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452
QY 72 --QSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGO-YGGNNAALVNQ 128
DB 453 IIEIANNKATDVGGGA-----YVKGLTTCNSHRLQFLKNSDKQGGIYGEDNITLSNL 507
QY 129 T 129
DB 508 T 508

RESULT 14
US-09-841-132-594
; Sequence 594, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 594
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: C. Trachomatis D serovar
US-09-841-132-594

Query Match      11.7%; Score 91; DB 9; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SGSALAGVVPWGCGGN--HNGGNSGSGPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
DB 394 NADAWASSPQSGGATTVNSGDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452
QY 72 --QSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGO-YGGNNAALVNQ 128
DB 453 IIEIANNKATDVGGGA-----YVKGLTTCNSHRLQFLKNSDKQGGIYGEDNITLSNL 507
QY 129 T 129
DB 508 T 508
```

RESULT 15

US-10-467-534-45
; Sequence 45, Application US/10467534
; Publication No. US20040131625A1
; GENERAL INFORMATION:
; APPLICANT: Berthet, Francois-Xavier Jacques
; APPLICANT: Lobet, Yves
; APPLICANT: Poolman, Jan
; APPLICANT: Verlant, Vincent Georges Christian Louis
; TITLE OF INVENTION: Vaccine Composition
; FILE REFERENCE: B45261
; CURRENT APPLICATION NUMBER: US/10/467,534
; CURRENT FILING DATE: 2003-08-08
; PRIOR APPLICATION NUMBER: PCT/EP02/01356
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: GB 0103169.9
; PRIOR FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-467-534-45

Query Match 11.7%; Score 91; DB 16; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SGSALAGVVPQWGGGN--HNGGGSSGPDSTLSIYQGSANAALALOSDARKSETTIT- 71
DB 394 NADAWASSPQSGSGATTVNSGDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452

QY 72 --QSGYGNADYQGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVQG- YGNNNAALVQ 128
DB 453 IIEIANNKATDVGGGA-----YVKGLTTCENSHRLQFLKNSSDKQGGIYGEDNITLSNL 507

QY 129 T 129
DB 508 T 508

Search completed: August 2, 2004, 15:36:12
Job time : 37.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-22
Perfect score: 776
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues
Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pep.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pep.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
|------------|-------|-------|--------|----|-------------|

| | | | | | | |
|----|-------|-------|------|----|----------------------|-------------------|
| 1 | 776 | 100.0 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 2 | 712 | 91.8 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 3 | 682 | 87.9 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 4 | 677 | 87.2 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 5 | 657 | 84.7 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 6 | 639 | 79.8 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 7 | 609 | 78.5 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 8 | 604 | 77.8 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 9 | 602 | 77.6 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 10 | 601 | 77.4 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 11 | 600 | 77.3 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 12 | 595 | 76.7 | 131 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 13 | 587 | 73.1 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 521 | 67.1 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 518 | 66.8 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 518 | 66.8 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 516 | 66.5 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 516 | 66.5 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 497 | 64.0 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 485 | 59.9 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 21 | 465 | 59.9 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 465 | 59.9 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 23 | 463 | 59.7 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 24 | 440 | 56.7 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 440 | 56.7 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 336 | 43.3 | 109 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 27 | 256 | 33.0 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 28 | 215 | 27.7 | 48 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 29 | 159.5 | 20.6 | 70 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 30 | 104.5 | 13.5 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 31 | 104.5 | 13.5 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 104.5 | 13.5 | 186 | 30 | US-10-417-886-5833 | Sequence 5833, Ap |
| 33 | 97.5 | 12.6 | 520 | 1 | PCT-US02-18256-21 | Sequence 21, Appl |
| 34 | 96 | 12.4 | 2308 | 33 | US-60-161-932-688 | Sequence 688, App |
| 35 | 96 | 12.4 | 2309 | 20 | US-09-614-150-25488 | Sequence 25488, A |
| 36 | 96 | 12.4 | 2309 | 20 | US-09-614-150A-25488 | Sequence 25488, A |
| 37 | 96 | 12.4 | 2309 | 33 | US-60-191-637-25607 | Sequence 25607, A |
| 38 | 96 | 12.4 | 2309 | 33 | US-60-191-681-20216 | Sequence 20216, A |
| 39 | 95.5 | 12.3 | 252 | 1 | PCT-US03-13414-24 | Sequence 24, Appl |
| 40 | 95 | 12.2 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 41 | 95 | 12.2 | 445 | 33 | US-60-360-039-20638 | Sequence 20638, A |
| 42 | 95 | 12.2 | 597 | 1 | PCT-US01-05992-146 | Sequence 146, App |
| 43 | 95 | 12.2 | 597 | 22 | US-09-793-305-146 | Sequence 146, App |
| 44 | 94 | 12.1 | 271 | 30 | US-10-437-963-147343 | Sequence 147343, |
| 45 | 93.5 | 12.0 | 145 | 21 | US-09-739-449-8854 | Sequence 8854, Ap |

ALIGNMENTS

RESULT 1
US-09-543-407-22
; Sequence 22 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; CURRENT APPLICATION NUMBER: US/09/543,407
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-22

Query Match 100.0%; Score 776; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.2e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-28
Sequence 28, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-28

Query Match 91.8%; Score 712; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 9.1e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDI 113
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDI 113

Qy 114 TVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 114 TVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5
Sequence 5, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRT
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 87.9%; Score 682; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 1.5e-64;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSITLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-08-233-642A-57
Sequence 57, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
TITLE OF INVENTION: BASED VACCINES
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 57
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 87.2%; Score 677; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 5.2e-64;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNYDQVTRVVTHEMAHADQNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNYDQVTRVVTHEMAHADQNAKNSDITVQYGG 120
QY 121 NNAALVNOTASDSSVMVROVGFQGNATANOY 151
Db 121 NNPALVNOTASDSSVMVROVGFQGNATANOY 151

RESULT 5
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30

Query Match 84.7%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 7.4e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADN-----YDQVTRVVTHEMAHADQ 105
Db 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQVTRVVTHEMAHA-- 118
QY 106 WNAKNSDITVGOYGGNNAALVNOTASDSSVMVROVGFQGNATANOY 151
Db 119 -----GGNNAALVNOTASDSSVMVROVGFQGNATANOY 151

RESULT 6
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20
Query Match 79.8%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 8.9e-58;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNYDQVTRVVTHEMAHA----- 103
Db 61 SDARK-----YDQVTRVVTHEMAHAGQGANSTIELTQNGFR 98
QY 104 -----DOWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROVGFQGNATANOY 151
Db 99 NNATIDOWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROVGFQGNATANOY 151
RESULT 7
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24

Query Match 78.5%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.1e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSSGPDSTLSIYQVGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGAD-----NYDQVTRVVT 97
Db 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNDYDQVTRVVT 120
QY 98 HEVAHADQWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROVGFQGNATANOY 151
Db 121 HEVAHA-----NQTASDSSVMVROVGFQGNATANOY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match          77.8%; Score 604; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.6e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYQTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match          77.6%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 6e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match          77.8%; Score 604; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.6e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYQTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match          77.6%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 6e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYGNAGDVGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
```

```
QY 121 NNAALVNYQTASDSSVMVROVGFNNATANQY 151
DB 121 LVTRVVTHEMAHASVMVROVGFNNATANQY 151

RESULT 10
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26

Query Match          77.4%; Score 601; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
DB 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNYQTASDSSVMVROVGFNNATANQY 151
DB 121 NNAALVNYQTASDSSVMVROVGFNNATANQY 151

RESULT 11
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18

Query Match          77.3%; Score 600; DB 19; Length 151;
```



```
/ EARLIER APPLICATION NUMBER: US 07/347,189
/ EARLIER FILING DATE: 1989-05-04
/ EARLIER APPLICATION NUMBER: US 07/789,437
/ EARLIER FILING DATE: 1991-11-06
/ EARLIER APPLICATION NUMBER: US 07/970,846
/ EARLIER FILING DATE: 1992-11-03
/ EARLIER APPLICATION NUMBER: US 08/187,865
/ EARLIER FILING DATE: 1994-01-28
/ EARLIER APPLICATION NUMBER: US 08/318,519
/ EARLIER FILING DATE: 1994-10-05
/ EARLIER APPLICATION NUMBER: US 08/495,959
/ EARLIER FILING DATE: 1995-06-28
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      56.8%; Score 518; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.3e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGGPDSTLSIYQYGSANAALAQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1 MKLLKVAATAIVFSGSAGVVPQYGGGNGHGGNSGGPNSLNIIYQGGNSALALQ 60
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      61 SDARKSETTITQSGYNGADVCGGADNVDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db      61 TPARNSDLTITQGGGNGADVCGGSDSSIDLTRGFNSATLTDWNGKNSMTVKQFGG 120
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      121 NNAALVNOTASDSSVMVROVGFNNATANOY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121 GNGAAVDQIASNSSVNVTVQVGFNNATAHOY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: August 2, 2004, 15:26:45
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pending Patents AA New.*

2: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*

3: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*

4: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*

5: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*

6: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*

7: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1 | 53.8 | 66.8 | 151 | US-09-741-873C-4 | Sequence 4, Appli |
| 2 | 44.0 | 56.7 | 131 | US-09-741-873C-2 | Sequence 2, Appli |
| 3 | 97.5 | 12.6 | 520 | US-10-479-638-21 | Sequence 21, Appli |
| 4 | 92 | 11.9 | 841 | US-60-565-632-7906 | Sequence 7906, Ap |
| 5 | 92 | 11.9 | 841 | US-60-579-062-7906 | Sequence 7906, Ap |
| 6 | 91 | 11.7 | 1751 | US-10-498-327-1115 | Sequence 115, App |
| 7 | 91 | 11.7 | 1751 | US-10-872-155-445 | Sequence 445, App |
| 8 | 91 | 11.7 | 1751 | US-10-872-155-594 | Sequence 594, App |
| 9 | 90 | 11.6 | 1010 | PCT-US04-21432-402 | Sequence 402, App |
| 10 | 89.5 | 11.5 | 388 | US-09-248-796A-17306 | Sequence 17306, A |
| 11 | 88.5 | 11.4 | 193 | US-10-425-115-254240 | Sequence 254240, |
| 12 | 87 | 11.2 | 892 | US-09-952-267B-5 | Sequence 5, Appli |
| 13 | 87 | 11.2 | 892 | US-10-872-768-5 | Sequence 5, Appli |
| 14 | 87 | 11.2 | 892 | US-10-872-768-5 | Sequence 5, Appli |
| 15 | 86.5 | 11.1 | 125 | US-10-425-115-33419 | Sequence 33419, |
| 16 | 86.5 | 11.1 | 595 | US-10-872-153-337 | Sequence 337, App |
| 17 | 86.5 | 11.1 | 1752 | US-10-872-153-180 | Sequence 180, App |
| 18 | 86 | 11.1 | 201 | US-10-425-115-309662 | Sequence 309662, |
| 19 | 86 | 11.1 | 244 | US-10-854-439-84 | Sequence 84, Appli |
| 20 | 85 | 11.0 | 197 | US-10-425-115-304391 | Sequence 304391, |
| 21 | 85 | 11.0 | 215 | US-09-248-796A-20202 | Sequence 20202, A |
| 22 | 85 | 11.0 | 239 | US-10-425-115-343835 | Sequence 343835, |
| 23 | 85 | 11.0 | 295 | US-10-425-115-312468 | Sequence 312468, |
| 24 | 85 | 11.0 | 573 | US-60-565-632-7907 | Sequence 7907, Ap |
| 25 | 85 | 11.0 | 573 | US-60-579-062-7907 | Sequence 7907, Ap |
| 26 | 85 | 11.0 | 688 | US-60-581-351-11388 | Sequence 11388, A |

| | | | | | | |
|----|------|------|------|---|----------------------|--------------------|
| 27 | 84.5 | 10.9 | 282 | 6 | US-10-501-282-2420 | Sequence 2420, Ap |
| 28 | 84.5 | 10.9 | 303 | 6 | US-10-501-282-2422 | Sequence 2422, Ap |
| 29 | 84 | 10.8 | 256 | 6 | US-10-425-115-301334 | Sequence 301334, |
| 30 | 84 | 10.8 | 619 | 1 | PCT-US04-10229-38 | Sequence 38, Appli |
| 31 | 84 | 10.8 | 753 | 6 | US-10-170-205E-35514 | Sequence 35514, A |
| 32 | 84 | 10.8 | 753 | 6 | US-10-170-205E-35515 | Sequence 35515, A |
| 33 | 84 | 10.8 | 1627 | 6 | US-10-170-205E-16659 | Sequence 16659, A |
| 34 | 84 | 10.8 | 1905 | 1 | PCT-US04-09388-9 | Sequence 9, Appli |
| 35 | 83 | 10.7 | 586 | 1 | PCT-US03-24982A-317 | Sequence 317, App |
| 36 | 82.5 | 10.6 | 286 | 6 | US-10-854-439-511 | Sequence 511, App |
| 37 | 82.5 | 10.6 | 687 | 6 | US-10-425-115-278239 | Sequence 278239, |
| 38 | 82 | 10.6 | 374 | 1 | PCT-US04-11210-37 | Sequence 37, Appli |
| 39 | 81.5 | 10.5 | 508 | 6 | US-10-425-115-285216 | Sequence 285216, |
| 40 | 81.5 | 10.5 | 532 | 6 | US-10-425-115-285214 | Sequence 18, Appli |
| 41 | 81.5 | 10.5 | 669 | 6 | US-10-489-425-18 | Sequence 7905, Ap |
| 42 | 81 | 10.4 | 412 | 7 | US-60-565-632-7905 | Sequence 7905, Ap |
| 43 | 81 | 10.4 | 412 | 7 | US-60-579-062-7905 | Sequence 11, Appli |
| 44 | 81 | 10.4 | 1358 | 6 | US-10-778-804-11 | Sequence 45603, A |
| 45 | 80.5 | 10.4 | 234 | 6 | US-10-767-701-45603 | |

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query March 66.8% Score 518; DB 5; Length 151;
Best Local Similarity 66.9%; Pred.No. 3.4e-37;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSITYQVGSNAALALQ 60
1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSITYQVGSNAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSITYQVGSNAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNDYDQIVTRVTHEVAHADOWNKAKSDITVGYGG 120
61 TDAANSLLTITQSGYNGADVGQGDNDYDQIVTRVTHEVAHADOWNKAKSEMIVTKVFGG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 GNGRAVDQTASNSSVNVTVQVGFNNATAHQY 151


```

; APPLICANT: Wu, Wei Bei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; TITLE OF INVENTION: Compositions thereof
; FILE REFERENCE: 38-21 (53403) C
; CURRENT APPLICATION NUMBER: US/60/579,062
; CURRENT FILING DATE: 2004-06-11
; NUMBER OF SEQ ID NOS: 41445
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7506
; LENGTH: 841
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (810)..(810)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; US-60-579-062-7906

```

| | | | | |
|-----------------------|--------|--|----------------|--------------------|
| Query Match | 11.9%; | Score 92; | DB 7; | Length 841; |
| Best Local Similarity | 28.3%; | Pred. No. 8.9; | | |
| Matches | 36; | Conservative 16; | Mismatches 59; | Indels 16; Gaps 5; |
| Qy | 30 | GNHGGG--NSGPPDSTLSIYQGSAAALALQADARKSETTITQSG-----YNGGADV | 81 | |
| Db | 657 | GNNGSAEATNGNADSQDQ-GSAATA-----NADNNANTDAQDAGQCHNGSGAAEI | 710 | |
| Qy | 82 | GCGADNYQLQVTRVTHEMAHDOWNAKNSDITVGYCGNNAALVNQTASDSVMVRQVG | 141 | |
| Db | 711 | NGGADGQNE--NGAAASSGNADQNQNNNSNNDSGNNNGNSVEENSKEDSGNESQEN | 768 | |
| Qy | 142 | FGNNATA | 148 | |
| Db | 769 | KDNNDTIS | 775 | |

```

RESULT 6
US-10-498-327-115
; Sequence 115, Application US/10498327
; GENERAL INFORMATION:
; APPLICANT: Grandi, Guido
; APPLICANT: Ratti, Giulio
; TITLE OF INVENTION: Immunisation Against Chlamydia Trachomatis
; FILE REFERENCE: 002441.00085
; CURRENT APPLICATION NUMBER: US/10/498,327
; CURRENT FILING DATE: 2004-06-10
; PRIOR APPLICATION NUMBER: PCT/IB02/05761
; PRIOR FILING DATE: 2002-12-12
; NUMBER OF SEQ ID NOS: 262
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 115
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-498-327-115

```

| | Query Match | 11.7% | Score 91 | DB 6 | Length 1751 |
|----|-----------------------|--|-----------------|---------------|-------------|
| | Best Local Similarity | 33.1% | Pred. No. 25 | | |
| | Matches | 40 | Conservative 10 | Mismatches 59 | Indels 12 |
| | | | | Gaps 5 | |
| Qy | 15 | SGSALAGVVPQGGGN--HNGGNSGPGDSTLSIYQYGSANAALQSDARKSETTIT- 71 | : | : | : |
| Db | 394 | NADAWASSPQSGGATVNSGDSGSSGSDTSETVPATAKG-GLYTDKNLSITITG 452 | : | : | : |
| Qy | 72 | --QSGYGNAGVGGCADNYDQLVTRVVFHEAHADQWNAKNSDITVGO-YGNNAAALVNQ 128 | : | : | : |
| Db | 453 | IIETANNKATDVGGGA----YVKGTLTCENSHRLQFLKNSSDKGGGIYGEDNITLSNL 507 | : | : | : |
| Qy | 129 | T 129 | | | |
| Db | 508 | T 508 | | | |

```

RESULT 7
US-10-872-155-445
; Sequence 445, Application US/10872155
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Prost, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C10
; CURRENT APPLICATION NUMBER: US/10/872,155
; CURRENT FILING DATE: 2004-06-18
; PRIOR APPLICATION NUMBER: 09/620,412
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 09/598,419
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/556,877
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 09/454,684
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 09/426,571
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 09/410,568
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/288,594
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/208,277
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FASTSEQ for Windows Version 3.0/4.0
; SEQ ID NO 445
; LENGTH: 1751
; TYPE: PR1
; ORGANISM: Chlamydia trachomatis serovar D
US-10-872-155-445

```

```

Query Match      11.7%; Score 91; DB 6; Length 1751;
Best Local Similarity 33.1%; Pred.No. 25;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY      15  S G S A L A G V P Q W G G G G N - - H N G G N S S G P D S T L S I Y Q Y G S A N A A L Q S D A R K S E T T I T - 71
DB      394  N A D A W A S S S F Q S G S G A T V S N G D S S G S D S D T S E T V P A T A K G G - G L Y T D K N L S I T N I T G 452

QY      72  - - Q S G Y N G A D V G G C A D N Y Q L Y T R V V T H E M A H A D O W N A K N S D I T V C Q - Y G G N N A A L V N Q 128
DB      453  I I E I A N N A K A T D V G G G A - - - - Y V K G T L T C E N S H R L Q F L K N S D Q K G G G I Y G E D N I T L S N L 507

QY      129  T 129
DB      508  T 508

RESULT 8
US-10-872-155-594
; Sequence 594, Application US/10872155
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C10
; CURRENT APPLICATION NUMBER: US/10/872,155
; CURRENT FILING DATE: 2004-06-18
; PSIOR APPLICATION NUMBER: 09/620,412
; PRIOR FILING DATE: 2000-07-20
; PSIOR APPLICATION NUMBER: 09/598,419
; PRIOR FILING DATE: 2000-06-20
; PSIOR APPLICATION NUMBER: 09/556,877
; PRIOR FILING DATE: 2000-04-19
; PSIOR APPLICATION NUMBER: 09/454,684
; PRIOR FILING DATE: 1999-12-03

```

```
; PRIOR APPLICATION NUMBER: 09/426,571
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 09/410,568
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 09/288,594
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 09/208,277
; PRIOR FILING DATE: 1998-12-08
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 594
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: C. Trachomatis D serovar
US-10-872-155-594

Query Match      11.7%; Score 91; DB 6; Length 1751;
Best Local Similarity 33.1%; Pred. No. 25;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SCSALAGVVPQGGGN--HNGCGNNGSGDPSTLSIYQYGSANAALQSDARKSETTIT- 71
Db 394 NADAWASSPQSGGATTVNSGDSGSDSTSTVPATKGG-GLYTDKNLSITNITG 452
QY 72 --QSGYGNGADYGGQADNYDQLVTRVVTHEMAHADQNAKNSDITVQY- YGNNAAALVNQ 128
Db 453 IIEIANNKATDVGGGA-----YVKGLTCENSHRLQFLKNSDDKQGGIYGEDNITLSNL 507
QY 129 T 129
Db 508 T 508

RESULT 9
PCT-US04-21492-402
; Sequence 402, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 402
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(30)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (393)...(428)
; OTHER INFORMATION: Carbohydrate binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (39)...(300)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (610)...(959)
; OTHER INFORMATION: Glycosyl hydrolases family 6
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (493)...(521)
```

```
; OTHER INFORMATION: Cellulose or protein binding domain
PCT-US04-21492-402

Query Match      11.6%; Score 90; DB 1; Length 1010;
Best Local Similarity 22.4%; Pred. No. 16;
Matches 32; Conservative 27; Mismatches 56; Indels 28; Gaps 5;

QY 12 IVVSGSALAGVVPQGW---GGNHNGGNGSGDPSTLSIYQYGSANAALQSDARKSET 68
Db 316 LTASGTLVKNIVSNWGTITGNGSSSSSSSSSSSS-----SSSSSSSSSSSSSSSS 367
QY 69 TITQSGYNGCA-----DVGGADNYDQLVTRVVTHEMAHADQW---NAKNSD-- 112
Db 368 SSGSTGGNCAGVNVYPNWTARDWSGAINHANAGDQVQNSLYRANVTNSVPGSDAS 427
QY 113 -ITVQYGGNNAALVNQTSADSS 134
Db 428 WTSIGACGGNGSTTSSSSSSSS 450

RESULT 10
US-09-248-796A-17306
; Sequence 17306, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17306
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17306

Query Match      11.5%; Score 89.5; DB 5; Length 388;
Best Local Similarity 22.1%; Pred. No. 6;
Matches 33; Conservative 17; Mismatches 44; Indels 55; Gaps 4;

QY 30 GNHNGGNGSGDPSTLSIYQYGSANAALQSDARKSETTITQSGYGGNGADYGGQADNYD 89
Db 132 GNQVGSNSYSPDT-----YGSAIGTLGVQEKTPAVTGIHSGIGAAA-----YP 177
QY 90 QLVTRVVTHEMAHADQNAKNSDITVQYGGNNAALVNQTS----- 131
Db 178 ELT-----NAGTGLAKGTAPASTATYGESPSADYSKSGATGVVPTATYLN 223
QY 132 -----DSSVMVQVGFQGNATANQY 151
Db 224 TSGAPTGSINTAGVGGAGFGDNTSSY 252

RESULT 11
US-10-425-115-254240
; Sequence 254240, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 254240
```



```
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_163446C.1.pep
US-10-425-115-254240

Query Match      11.4%; Score 88.5; DB 6; Length 193;
Best Local Similarity 26.3%; Pred. No. 3.3;
Matches 44; Conservative 22; Mismatches 62; Indels 39; Gaps 8;

QY 5 KYAFAAIVVSGSALAGV-----PWWGGGNNHGGNSSGP-----DSTLS 46
Db 5 KLISGLIVLIGIGLANAVRVARYSADGTGTGEGCGGYVNGGSGSGSGTGGSDS-- 62
QY 47 IYQYGSANAALALQSDARKSETTITQSGYNGADVGOGADNYDOLVTRVVTHEMAHADQW 106
Db 63 --PYGAHASAGGGAGGASQN--GGSGYGGSGSGSGSSTYSQ-----GGYISGYG 110
QY 107 NAKNSDITVGYGGNNA-ALVNTASDSSVMVRQVFGNNAT-ANQY 151
Db 111 ESSNAGGTGGGGGQAGGYWNSNAQSG-----SGTGGSGSYANRY 152

RESULT 12
US-09-952-267B-5
; Sequence 5; Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267B-5

Query Match      11.2%; Score 87; DB 5; Length 892;
Best Local Similarity 26.8%; Pred. No. 26;
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNHN-----GGNNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGY--- 75
Db 89 GGGDYNEAKGNYSTVGGSSNTAKGEKSTIGGGDTN-----DANGTYSTIGGGYSRA 141

QY 76 -GNGADVGGADNYDOLVTRVVTHEMAHADOWNNAKNSDITVQYG---GNNALV---N 127
Db 142 IGSSTIGGG--YNNQATGEKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSHN 199

QY 128 QTASDSSVMVRQVFGNNATN 149
Db 200 QATGEGSF---AAGVENKANAN 218

RESULT 13
US-09-952-267B-5
; Sequence 5; Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/10/872,769
; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-769-5

Query Match      11.2%; Score 87; DB 6; Length 892;
Best Local Similarity 26.8%; Pred. No. 26;
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNHN-----GGNNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGY--- 75
Db 89 GGGDYNEAKGNYSTVGGSSNTAKGEKSTIGGGDTN-----DANGTYSTIGGGYSRA 141

QY 76 -GNGADVGGADNYDOLVTRVVTHEMAHADOWNNAKNSDITVQYG---GNNALV---N 127
Db 142 IGSSTIGGG--YNNQATGEKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSHN 199

QY 128 QTASDSSVMVRQVFGNNATN 149
Db 200 QATGEGSF---AAGVENKANAN 218

RESULT 14
US-10-872-769-5
; Sequence 5; Application US/10872769
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/10/872,769
; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-769-5

Query Match      11.2%; Score 87; DB 6; Length 892;
Best Local Similarity 26.8%; Pred. No. 26;
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNHN-----GGNNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGY--- 75
Db 89 GGGDYNEAKGNYSTVGGSSNTAKGEKSTIGGGDTN-----DANGTYSTIGGGYSRA 141

QY 76 -GNGADVGGADNYDOLVTRVVTHEMAHADOWNNAKNSDITVQYG---GNNALV---N 127
Db 142 IGSSTIGGG--YNNQATGEKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSHN 199

QY 128 QTASDSSVMVRQVFGNNATN 149
Db 200 QATGEGSF---AAGVENKANAN 218
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|--------|--------------------|
| 1 | 682 | 87.9 | 151 | 2 | JC6039 | fimbrin protein ag |
| 2 | 682 | 87.9 | 151 | 2 | A10635 | major curlin chain |
| 3 | 521 | 67.1 | 151 | 2 | S70788 | curlin protein csq |
| 4 | 499.5 | 64.4 | 152 | 2 | D90806 | curlin major subun |
| 5 | 499.5 | 64.4 | 152 | 2 | H85665 | hypothetical prote |
| 6 | 109 | 14.0 | 1748 | 2 | S42136 | cnjB protein - Tet |
| 7 | 109 | 14.0 | 2174 | 2 | E95965 | hypothetical glyci |
| 8 | 102 | 13.1 | 586 | 2 | T26667 | hypothetical prote |
| 9 | 98 | 12.6 | 615 | 2 | E70663 | probable pPE prote |
| 10 | 95 | 12.2 | 590 | 2 | E70946 | probable pPE prote |
| 11 | 94 | 12.1 | 407 | 2 | T21956 | hypothetical prote |
| 12 | 93.5 | 12.0 | 145 | 2 | AD3143 | conserved hypotet |
| 13 | 93.5 | 12.0 | 145 | 2 | H98144 | hypothetical prote |
| 14 | 93.5 | 12.0 | 645 | 2 | F70825 | probable pPE prote |
| 15 | 93.5 | 12.0 | 1567 | 2 | S11672 | ice nucleation pro |
| 16 | 92.5 | 11.9 | 151 | 2 | S70787 | curlin nucleator p |
| 17 | 92.5 | 11.9 | 151 | 2 | C90806 | minor curlin subun |
| 18 | 92.5 | 11.9 | 151 | 2 | G85665 | curlin minor chain |
| 19 | 91.5 | 11.8 | 1034 | 2 | JC2143 | ice nucleation act |
| 20 | 91.5 | 11.8 | 1258 | 2 | JQ0188 | ice nucleation pro |
| 21 | 91 | 11.7 | 1751 | 2 | G71518 | hypothetical prote |
| 22 | 90.5 | 11.7 | 151 | 2 | JC6040 | fimbrin protein ag |
| 23 | 90.5 | 11.7 | 151 | 2 | AH0635 | nucleation compone |
| 24 | 90.5 | 11.7 | 1053 | 2 | B70987 | probable pPE prote |
| 25 | 89.5 | 11.5 | 1322 | 2 | S07053 | ice nucleation pro |
| 26 | 89 | 11.5 | 575 | 2 | S35327 | protein kinase egg |
| 27 | 87 | 11.2 | 251 | 2 | D95010 | hypothetical expor |
| 28 | 87 | 11.2 | 434 | 2 | E70768 | hypothetical glyci |
| 29 | 87 | 11.2 | 573 | 2 | C86266 | F3F19.21 protein - |

```
30      87  11.2  590  1  A45621  leishmanolysin (EC
31      87  11.2  599  2  B42049  leishmanolysin (EC
32      87  11.2  599  2  A44951  leishmanolysin (EC
33      87  11.2  602  1  P40221  leishmanolysin (EC
34      87  11.2  646  1  S19916  leishmanolysin (EC
35      86.5  11.1  262  2  S00275  tail fiber protein
36      86.5  11.1  487  2  C70830  probable pPE prote
37      86.5  11.1  1028  2  A56038  DNA-binding protei
38      86.5  11.1  1213  2  S6356  ovo protein 2, frui
39      86  11.1  582  2  F70675  probable pPE prote
40      85.5  11.0  940  2  D89723  protein F39D8.1b (
41      85.5  11.0  945  2  T21998  hypothetical prote
42      85.5  11.0  1210  2  A25547  ice nucleation pro
43      85  11.0  354  2  B70663  probable pPE prote
44      85  11.0  1052  2  A72959  conserved hypotet
45      85  11.0  1341  2  H98323  hypothetical prote
```

ALIGNMENTS

RESULT 1

```
JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 179, 662-667, 1996
A/Ritle: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.
A/Reference number: JC6039; MUID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:gl1184712; PIDN:AAC43599.1; PID:gl1184714
A/Accession: PC6015
A/Molecule type: protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Ritle: Purification and characterization of thin, aggregative fimbrinae from Salmonella
A/Reference number: A44898; MUID:91310586; PMID:1677357
A/Contents: 27655
A/Accession: A44898
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBIP:45936)
C/Genetics:
A/Gene: agfa
C/Function:
A/Description: major component of thin aggregative fimbrinae
A/Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbrina
F/1-20/Domain: signal sequence #status predicted <SIG>
F/21-151/Product: fimbrin protein agfa #status experimental <MAT>
```

Query Match 87.9%; Score 682; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 1.2e-50;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQVGSGNGNNGSGSPDSTLSIYQYGSANALALQ 60

1 MKLLKVAFAAIVVSGSALAGVVPQVGSGNGNNGSGSPDSTLSIYQYGSANALALQ 60

61 SDARKSETTITQSGYNGADVGCGADNYQLVTRVVTTHMAHADONAKNSDITTVQYGG 120

61 SDARKSETTITQSGYNGADVGCGADNYQLVTRVVTTHMAHADONAKNSDITTVQYGG 120

121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Garra, P.

Nature 413, 848-852, 2001

A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Garra, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 87.9%; Score 682; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 1.2e-50;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNSGDPSTLSIYQGSANAALALQ 60

DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNSGDPSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and CsgA

A:Reference number: S70783; MUID:96414458; PMID:8817489

A:Accession: S70788

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:gl147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cs

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:gl787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Eps sigma factor relieves H-NS-mediated transcriptional repression of csgA

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V',8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RQDSGLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 521; DB 2; Length 151;

Best Local Similarity 69.5%; Pred. No. 4.6e-37;

Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNSGDPSTLSIYQGSANAALALQ 60

DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNSGDPSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 120

DB 61 TDARNSDLTITQGGGNGADVGGDDSSIDITQGFNGSATLDOWNGKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: D90806

R:Hayaishi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kunata, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:gl3360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 64.4%; Score 499.5; DB 2; Length 152;

Best Local Similarity 67.8%; Pred. No. 3e-35; 30; Indels 1; Gaps 1;

Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNSGDPSTLSIYQGSANAALALQ 59

DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNSGDPSTLSIYQGSANAALALQ 60

QY 60 QSDARKSETTITQSGYNGGADVGGADNYDQVLTVRVVTHEMAHADOWNAKNSDITVQYGG 119


```

Db      294 SD---NDTVTNNYNDNNGNHNSS--SSNAITDAPGNAGDQQSQSD----- 341
QY      121 NNAALVNQTASD 132
Db      342 NNYGNANQDSYD 353

RESULT 12
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3143
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: PIDN:AAU45562.1; PID:g17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

Query Match      12.0%; Score 93.5; DB 2; Length 145;
Best Local Similarity 21.8%; Pred. No. 0.48;
Matches 34; Conservative 25; Mismatches 64; Indels 33; Gaps 4;

QY      3 LLKVAFAAIYVSGSALAGVVPQWG-----GGNHNGGNSGSPDSTLSIYQYGANA 55
Db      1 MIRKSFIAALVALVGLSAAAFAMANDVRIEQYGSNSAGGAEQYGNRIYQNGYNR 60
QY      56 ALALQSDARKSETTITQSGYNGADVGO-GADNYDQLVTRVVTHEMAHADQWNAKNSDIT 114
Db      61 IVG-----HQYGRNLSAVQEGHDNYGSTTQ-----NGNRNVAG 95

RESULT 13
H98144
hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98144
R:Goodner, B.; Hinkle, G.; Gattung, S.; Muller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98144
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:g15158413; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_228
A:Map position: linear chromosome

Query Match      12.0%; Score 93.5; DB 2; Length 145;
Best Local Similarity 21.8%; Pred. No. 0.48;
Matches 34; Conservative 25; Mismatches 64; Indels 33; Gaps 4;

QY      3 LLKVAFAAIYVSGSALAGVVPQWG-----GGNHNGGNSGSPDSTLSIYQYGANA 55

```

```

Db      1 MIRKSFIAALVALVGLSAAAFAMANDVRIEQYGSNSAGGAEQYGNRIYQNGYNR 60
QY      56 ALALQSDARKSETTITQSGYNGADVGO-GADNYDQLVTRVVTHEMAHADQWNAKNSDIT 114
Db      61 IVG-----HQYGRNLSAVQEGHDNYGSTTQ-----NGNRNVAG 95

RESULT 14
F70825
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70825
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70825
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-645 <COL>
A:Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17522.1; PID:e12532
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match      12.0%; Score 93.5; DB 2; Length 645;
Best Local Similarity 24.5%; Pred. No. 2.6;
Matches 39; Conservative 17; Mismatches 68; Indels 35; Gaps 6;

QY      15 SGSALAGVVPQ-----WGGGNGHNGGNSGSPDSTLSIYQYGANAALQSDARKSETT 69
Db      237 SGNVNGVIGPSSFNVSGNIGNANVGGNSG-DNNFPGNFGNANIGNAGPNMSSPAV 295
QY      70 ITQS-----GYNGADVGOADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYVG----- 119
Db      296 PTPNGNVGIGNGNGNFGGNTGN-----ANIGLVNGVNGVNGVNGSGSYNFGF 345
QY      120 ---GNNALVNQTASDS-----SYMVRQVGFNNATAN 149
Db      346 GNTGNNNIGLITGSIQIGFGLNSGSGNIGFNGSGTGN 384

RESULT 15
S11672
ice nucleation protein - Xanthomonas campestris
C:Species: Xanthomonas campestris
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
C:Accession: S11672
R:Zhao, J.; Orser, C.S.
Mol. Gen. Genet. 223, 163-166, 1990
A:Title: Conserved repetition in the ice nucleation gene inax from Xanthomonas campestr
A:Reference number: S11672; MUID:91080859; PMID:2259339
A:Accession: S11672
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1567 <ZHA>
A:Cross-references: EMBL:X52970; NID:g48531; PIDN:CAA37140.1; PID:g48532
C:Superfamily: ice nucleation protein

Query Match      12.0%; Score 93.5; DB 2; Length 1567;
Best Local Similarity 28.2%; Pred. No. 6.9;
Matches 42; Conservative 21; Mismatches 41; Indels 45; Gaps 10;

QY      34 GCGNSSGPDSTLSIYQYGANAALQSDARKSETTITQSGYNGADVQGGD---NY-- 88

```

```

Db      383 GSTGTAGADSTL-IAGYGSTQTA-----GGESSLT-AGYGTQTARQGSDDITAGYGS 432
QY      89 -----DQVTRVVTHEMAHAD-----QWNAKNSDITVQYQ-----GNNAAALV 126
Db      433 TGTAGADSTLIAGYGSTQTSQSDSSLTAGYGSTQTARKGSDITAG-YGSTGTAGSDSLI 491
QY      127 -----NOTASDSSVMVRQVFGNNATANQ 150
Db      492 AGYGTQTAGSESSLT--AGYGTQTAAQ 518

```

Search completed: August 2, 2004, 14:56:24
 Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 682 | 87.9 | 151 | 1 | CSGA_SALTY |
| 2 | 521 | 67.1 | 151 | 1 | CSGA_ECOLI |
| 3 | 499.5 | 64.4 | 152 | 1 | CSGA_ECO57 |
| 4 | 93.5 | 12.0 | 1567 | 1 | ICEN_XANGT |
| 5 | 92.5 | 11.9 | 151 | 1 | CSGB_ECOLI |
| 6 | 91.5 | 11.8 | 1034 | 1 | ICEN_PANAN |
| 7 | 91.5 | 11.8 | 1258 | 1 | ICEN_ERWHE |
| 8 | 91 | 11.7 | 1754 | 1 | PMPB_CHLTR |
| 9 | 90.5 | 11.7 | 151 | 1 | CSGB_SALTY |
| 10 | 90.5 | 11.7 | 151 | 1 | CSGB_PANAN |
| 11 | 89.5 | 11.5 | 1322 | 1 | ICEN_PANAN |
| 12 | 87 | 11.2 | 491 | 1 | YK96_MYCTU |
| 13 | 87 | 11.2 | 590 | 1 | GP63_LEIDO |
| 14 | 87 | 11.2 | 599 | 1 | GP63_LEICH |
| 15 | 87 | 11.2 | 602 | 1 | GP63_LEIVA |
| 16 | 87 | 11.2 | 646 | 1 | GP63_LEIME |
| 17 | 86.5 | 11.1 | 262 | 1 | V338_BPT2 |
| 18 | 86.5 | 11.1 | 487 | 1 | Y442_MYCTU |
| 19 | 86.5 | 11.1 | 548 | 1 | CPAK_ECOLI |
| 20 | 86.5 | 11.0 | 1028 | 1 | OVO_DROME |
| 21 | 85.5 | 11.0 | 1210 | 1 | ICEN_PSEFL |
| 22 | 85 | 11.0 | 1778 | 1 | N189_SCHPO |
| 23 | 84.5 | 10.9 | 495 | 1 | P033_MOUSE |
| 24 | 84.5 | 10.9 | 760 | 1 | YBTL_ECOLI |
| 25 | 84.5 | 10.9 | 1211 | 1 | BUN2_DROME |
| 26 | 83.5 | 10.8 | 497 | 1 | P033_RAT |
| 27 | 83 | 10.7 | 1140 | 1 | YW96_YEAST |
| 28 | 82 | 10.6 | 500 | 1 | P033_HUMAN |
| 29 | 82 | 10.6 | 678 | 1 | YF48_MYCTU |
| 30 | 81.5 | 10.5 | 392 | 1 | HWE1_HUMAN |
| 31 | 81.5 | 10.5 | 737 | 1 | YD70_MYCFN |
| 32 | 81.5 | 10.5 | 2038 | 1 | FSH_DROME |
| 33 | 81 | 10.4 | 165 | 1 | GRPI_ORISA |

34 80.5 10.4 874 1 ALE6_AZOVI
35 80.5 10.4 1196 1 ICEV_PSEX
36 80 10.3 204 1 CO3A_MEDSA
37 80 10.3 362 1 ALBI_STACP
38 80 10.3 429 1 DR48_YEAST
39 80 10.3 443 1 YP98_MYCTU
40 80 10.3 543 1 YP91_MYCTU
41 79.5 10.2 396 1 PER_DROPV
42 78.5 10.1 346 1 RO21_XENLA
43 78.5 10.1 1148 1 ICEK_PSEX
44 78.5 10.1 1317 1 N145_YEAST
45 78 10.1 362 1 P35_MYCPE

ALIGNMENTS

RESULT 1
CSGA_SALTY STANDARD; PRT; 151 AA.
ID P5225;
AC P5225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Rowling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli" fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.;
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Bartell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531357; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kocoyanni V., Schwartz D.C., Blattner P.R.;

Q92fh0 azotobacter
Q33479 pseudomonas
Q07202 medicago sa
Q05156 staphylococ
P18999 saccharomyc
Q10340 mycobacteri
Q50630 mycobacteri
P91698 drosophila
P51989 xenopus lae
Q30611 pseudomonas
P49687 saccharomyc
Q50367 mycoplasma

```

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL J. Bacteriol. 185:2330-2337(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Cloughier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae.";
RN J. Bacteriol. 178:662-667(1996).
RN [6]
RP SEQUENCE OF 21-151 FROM N.A.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=94013373; PubMed=8104955;
RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae.";
RN J. Clin. Microbiol. 31:2263-2273(1993).
RN [7]
RP SEQUENCE OF 21-33.
RC SPECIES=S. enteritidis; STRAIN=27655-3B;
RX MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeedy L., Mueller K.M., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ002301; CA005317.1; -
DR EMBL; AF008749; RAL20074.1; -
DR EMBL; AL627269; CAD08268.1; -
DR EMBL; AF016840; AAC09399.1; -
DR EMBL; U43280; AAC43599.1; -
DR PIR; JC6039; JC6039.
DR StyGene; SGI0608; CSGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 134 151 SWMVQVGFNNATANQY -> DSVTQVAS (IN
FT REF. 6).
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
Query Match 87.9%; Score 682; DB 1; Length 151;
Best Local Similarity 90.7%; Pred. No. 2.2e-50;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHNGGNSGGPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGCGADNDYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120
DB 61 SPARKSETTITGSGYNGADVGCGADNDYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120
QY 121 NNAALVNQTPASDSSVMVQVGFNNATANQY 151
DB 121 NNAALVNQTPASDSSVMVQVGFNNATANQY 151

```

```

RESULT 2
CSGA_ECOLI STANDARD; PRT; 151 AA.
ID _CSGA_ECOLI
AC P28307;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR B1042.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=93211294; PubMed=8459772;
RA Olsen A., Arngvist A.;
RT "The Kps sigma factor relieves H-NS-mediated transcriptional
repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli.";
RN Mol. Microbiol. 7:523-536(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414488; PubMed=8817489;
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
fibronectin- and congo red-binding curli polymers in Escherichia coli
K-12.";
RN Mol. Microbiol. 18:661-670(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RN DNA Res. 3:137-155(1996).
RN [5]
RP SEQUENCE OF 21-40.
RC STRAIN=K12 / YNML;
RX MEDLINE=93023873; PubMed=1357528;
RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
RT "The Crl protein activates cryptic genes for curli formation and
fibronectin binding in Escherichia coli HB101.";
RN Mol. Microbiol. 6:2443-2452(1992).
RN [6]
RP SEQUENCE OF 21-31.
RC MEDLINE=91310586; PubMed=1677357;
RA Collinson S.K., Emeedy L., Trust T.J., Kay W.W.;
RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
RN J. Bacteriol. 173:4773-4781(1991).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.

```

```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04979; AAA23616.1; -
CC EMBL; X90754; CAA62282.1; -
CC EMBL; AE000205; AAC74126.1; -
CC EMBL; D90741; BAA35832.1; -
CC EMBL; D90742; BAA35840.1; -
CC PIR; S70788; S70788.
CC EcoGene; EG11489; csGA.
CC Fimbrin; Signal; Complete proteome.
FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
FT CHAIN 21 151
FT CONFLICT 7 A -> E (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 67.1%; Score 521; DB 1; Length 151;
Best Local Similarity 69.5%; Pred. No. 6e-37;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYGNAGDVGGADNYDQVTRVVTHEMAHADQWAKNSDITVGYGG 120
DB 61 TDARNSDLTITQHGNGGADVGGSDSSIDLTQGFNGSATLDQWNGKNSMTVKQFG 120

QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVTVQVGFNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DI 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR EC81420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csd promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2376(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posral G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature. 409:529-533(2001).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURL. CURL ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURL CAN BIND TO
CC FIBRONECTIN.
CC -----
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP275733; AAK53212.1; -
CC EMBL; AE005315; AAG55788.1; -
CC EMBL; AP002554; BAB34843.1; -
CC PIR; D90806; D90806.
CC PIR; H85665; H85665.
CC Fimbrin; Signal; Complete proteome.
KW SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT SEQUENCE 152 AA; 15099 MW; EE2D294DDE91243 CRC64;

Query Match 64.4%; Score 499.5; DB 1; Length 152;
Best Local Similarity 67.8%; Pred. No. 3.8e-35;
Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGNGHNGSGSPDSTLSIYQYGSANAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAL 60
QY 60 QSDARKSETTITQSGYGNAGDVGGADNYDQVTRVVTHEMAHADQWAKNSDITVGYGG 119
DB 61 QADARNSDLTITQHGNGGADVGGSDSSIDLTQGFNGSATLDQWNGKNSMTVKQFG 120

QY 120 GNNAAVNOTASDSSVMVQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVTVQVGFNNATAHQY 152

RESULT 4
ICEN_XANCT STANDARD; PRT; 1567 AA.
ID ICEN_XANCT
AC P18127;
DC 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X56S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens."
RL Mol. Gen. Genet. 223:163-166(1990).

```


QY 58 ALQSDARKSETTITGSGYGN-----GADVQCG-----ADNYDQLVTRVVTHMAHAD- 104
 Db 219 -----GESSSQWAGYSGTQGMKGDLTAGYSGTGTAGDSSSLIAGYSGTGTAGDS 270
 QY 105 -----QWNAKNSDITVQYGGNNAALVNTQATSSSVWVRQVGFNNATANO 150
 Db 271 SLTAGYGTQTAQKGSDLTAG-YGSTGTA-----GADSSLI-----AGYSGTGTAGE 316

RESULT 8

PMPB_CHLTR
 ID PMPB_CHLTR STANDARD; PRT; 1754 AA.
 AC O84418;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Probable outer membrane protein pmpB precursor (Polymorphic membrane protein B).
 DE PMPB OR C7413.
 GN Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/JW-3/Cx;
 RC MEDLINE=99000809; PubMed=9784136;
 RX Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
 RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";
 RT Science 282:754-759(1998).
 RL -!- SUBCELLULAR LOCATION: Cell wall surface (elementary bodies) (potential).
 CC -!- SIMILARITY: Belongs to the PMP outer membrane protein family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB001314; AAC68010.1; ALT_INIT.
 DR PHC1-2DPAGE; O84418; .
 DR InterPro; IPR003368; Chlamydia_PMP.
 DR Pfam; PF02415; Chlamydia_PMP; 4.
 DR TIGRFAWS; TIGR01376; POMP-repeat; 18.
 KW Outer membrane; Signal; Multigene family; Complete proteome.
 FT SIGNAL 1 14 POTENTIAL.
 FT CHAIN 15 1754 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
 SQ SEQUENCE 1754 AA; 183317 MW; 969CF8D85D36185D CRC64;

Query Match 11.7%; Score 91; DB 1; Length 1754;
 Best Local Similarity 33.1%; Pred. No. 7.3;
 Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;
 QY 15 GSAALAGVVPQWGGGN--HNGGNSGQPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
 Db 397 NADAWASSPQSGGATTVNSGDSGSDSDTSETVPATKGG-GLYTDKNLSITNITG 455
 QY 72 --QSGYNGADVQGANVDQLVTRVTHMAHADQWNAKNSDITVQY-YGNNNAALVNO 128
 Db 456 IIEIANNKATDVGGA-----YVXGTLTCNSHRLQFLKNSDQKGGIYGEDNITLSNL 510
 QY 129 T 129
 Db 511 T 511

RESULT 9

CSGB_SALTY STANDARD; PRT; 151 AA.

ID CSGB_SALTY STANDARD; PRT; 151 AA.
 AC Q827M3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor curlin subunit precursor.
 DE CSGB OR STY1180 OR T1777.
 GN Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RC MEDLINE=21534947; PubMed=11677608;
 RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Hague A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whithead S., Barrrell B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18.";
 RT Nature 413:848-852(2001).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL627269; CAD08267.1; .
 DR EMBL; AB016840; AAC69400.1; .
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326573495 CRC64;
 Query Match 11.7%; Score 90.5; DB 1; Length 151;
 Best Local Similarity 28.8%; Pred. No. 0.53;
 Matches 34; Conservative 19; Mismatches 46; Indels 19; Gaps 6;
 QY 51 GSAANAALALQSDARKSE-----TTTQSGYNGADVQY-GADNYDQL-----VTRVTH 98
 Db 14 GAGCIATATNYDLARSEYFNAVNELSKSFFNQAAIIGVGTGDNARSVRQEGSKLLSVISQ 73
 QY 99 E-----MAHADQWNAKNSDIT-VQYCGNNAALVNTQATSSSVWVRQVGFNNATANOY 151
 Db 74 EGENNPAKVDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAITQKSGNKNITQY 129

RESULT 10

CSGB_SALTY STANDARD; PRT; 151 AA.
 ID CSGB_SALTY

```

AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Pimbrin SEF17 minor subunit).
GN CSGS OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis,
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=9550497;
RA Collinson S.K., Plouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RL J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ002301; CAA05316.1; -
CC EMBL; AE008749; AAL20073.1; -
CC EMBL; U43280; AAC43598.1; -
CC FIR; JC6040; JC6040.
CC StyGene; SG10509; csGB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
FT SEQUENCE 151 AA; 16182 MW; CQPC543086DD361D CRC64;
SQ
Query Match 11.7%; Score 90.5; DB 1; Length 151;
Best Local Similarity 27.5%; Pred. No. 0.53;
Matches 36; Conservative 17; Mismatches 39; Indels 39; Gaps 5;
QY 10 AAIIVGSGALAGVGPQWGGGNGHGGNSGPDSTLSIYQYGSANAALQSDARKSETT 69
DB 58 ARVQEGSKLLSVISQ--EGGNRAKVDQAGNYFAYIEQIGNAN-----DAS 103
QY 70 ITQSGYGNAGDVQ--GADNYDQLVTRVVTHEMAHADQNAKNSDITVQYQGNNAALVQ 128

```

```

DB 104 ISQSAVNSAIIQKSGN-----KANIT--QYGTQKTAVVQ 139
QY 129 TASDSSVMVRQ 139
DB 140 KQSHVAIRVTQ 150
RESULT 11
ID ICEA_PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K.; Watabe S.; Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity.";
RL FEBS Lett. 258:297-300(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17316; CAA35194.1; -
CC PIR; S07053; S07053.
CC HSSP; P06620; 1INA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 69.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 49.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
FT SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;
SQ
Query Match 11.5%; Score 89.5; DB 1; Length 1322;
Best Local Similarity 26.8%; Pred. No. 7.1;
Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;
QY 34 GGNSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYGNAGDVQGGADNYDQLVT 93
DB 933 GSTSTAGPSSL-IAGYGSTQTA-----GYSILTAGY----- 965
QY 94 RVVTHEMAHADQNAKNSDITVQYQV-----GNNALV-----NQTSADSSVMVRQVGG 143
DB 966 -----STQTQGENSLDTTG-YGSTSTAGYESSLIAGYGSTQTASPFSKILM--AGYG 1013
QY 144 NNATANQ 150
DB 1014 SSQTARE 1020

```



```
FT METAL 255 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 BY SIMILARITY.
FT DISULFID 178 BY SIMILARITY.
FT DISULFID 301 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 565 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;

Query Match 11.2%; Score 87; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 4.7;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADNYDQLVTRVVTHEMAHA 103
DB 238 ASRYDQLVTRVVTHEMAHA 256

RESULT 14
GP63 LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence.";
RL Mol. Biochem. Parasitol. 39:267-274(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form.";
RL J. Biol. Chem. 267:1888-1895(1992).
CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-|-Leu-Lys-Iys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
EMBL; M80672; AAA29238.1; -.
```

```
DR EMBL; M88527; AAA29235.1; -.
DR PIR; A44951; A44951.
DR KSSP; P08148; ILML.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8_1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8B2A2E7C CRC64;

Query Match 11.2%; Score 87; DB 1; Length 599;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADNYDQLVTRVVTHEMAHA 103
DB 248 ASRYDQLVTRVVTHEMAHA 266

RESULT 15
GP63 LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Burton L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729(1988).
RN [2]
RP REVISIONS.
RA Burton L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589(1990).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A., Honans S.W., Bordier C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
```


FT STRAND 540 542
FT HELIX 543 545
FT TURN 546 546
FT STRAND 550 550
FT TURN 552 553
FT STRAND 555 557
FT HELIX 561 565
FT TURN 566 567
FT HELIX 569 572
FT TURN 573 573
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43B CRC64;
Query Match 11.2%; Score 87; DB 1; Length 602;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 85 ADNYDQLVTRVVTHEMAHA 103
Db 251 ASRYDQLVTRVVTHEMAHA 269

Search completed: August 2, 2004, 14:49:31
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-22
Perfect score: 776
Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mbc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rhodospirillum:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rviro:*
 - 16: sp_bacterioph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-----------|--------------------|
| 1 | 673 | 86.7 | 152 | 2 O33802 | O33802 salmonella |
| 2 | 566.5 | 73.0 | 150 | 2 Q7X243 | Q7X243 citrobacter |
| 3 | 538 | 69.3 | 149 | 2 Q7X240 | Q7X240 citrobacter |
| 4 | 496.5 | 64.0 | 152 | 16 Q8CW63 | Q8CW63 escherichia |
| 5 | 421.5 | 54.3 | 150 | 2 Q7X237 | Q7X237 enterobacte |
| 6 | 305.5 | 39.4 | 76 | 2 Q54069 | Q54069 salmonella |
| 7 | 122 | 15.7 | 29 | 2 Q983J5 | Q983J5 escherichia |
| 8 | 115 | 14.8 | 139 | 16 Q8EIH3 | Q8EIH3 shewanella |
| 9 | 110 | 14.2 | 130 | 16 Q8EJI4 | Q8EJI4 bradyrhizob |
| 10 | 109.5 | 14.1 | 502 | 16 Q8EIH4 | Q8EIH4 shewanella |
| 11 | 109 | 14.0 | 1748 | 5 Q94821 | Q94821 tetrahymena |
| 12 | 109 | 14.0 | 2174 | 16 Q92J08 | Q92J08 rhizobium m |
| 13 | 106.5 | 13.7 | 3501 | 16 Q8Y106 | Q8Y106 ralstonia s |
| 14 | 106.5 | 13.7 | 3552 | 16 Q8XSD6 | Q8XSD6 ralstonia s |
| 15 | 106 | 13.7 | 179 | 2 O33801 | O33801 salmonella |
| 16 | 105 | 13.5 | 1422 | 16 Q8EFU3 | Q8EFU3 shewanella |

| | | | | | |
|----|------|------|------|-----------|---------------------|
| 17 | 104 | 13.4 | 714 | 16 Q7U5X6 | Q7U5X6 synechococc |
| 18 | 102 | 13.1 | 191 | 10 Q7XDR3 | Q7XDR3 oryza sativ |
| 19 | 102 | 13.1 | 586 | 5 Q9NAJ4 | Q9NAJ4 caenorhabdi |
| 20 | 98 | 12.6 | 614 | 16 Q7TYR8 | Q7TYR8 mycobacteri |
| 21 | 98 | 12.6 | 615 | 16 P95249 | P95249 mycobacteri |
| 22 | 96.5 | 12.4 | 171 | 16 Q89J13 | Q89J13 bradyrhizob |
| 23 | 96 | 12.4 | 2310 | 5 Q9W2U7 | Q9W2U7 drosophila |
| 24 | 95.5 | 12.3 | 3859 | 16 Q98JN6 | Q98JN6 rhizobium 1 |
| 25 | 95 | 12.2 | 589 | 16 Q7TXS3 | Q7TXS3 mycobacteri |
| 26 | 95 | 12.2 | 590 | 16 Q53309 | Q53309 mycobacteri |
| 27 | 95 | 12.2 | 603 | 16 Q8VJ66 | Q8VJ66 mycobacteri |
| 28 | 94.5 | 12.2 | 157 | 16 Q88HG0 | Q88HG0 pseudomonas |
| 29 | 94.5 | 12.2 | 738 | 5 O02402 | O02402 pinctada fu |
| 30 | 94.5 | 12.2 | 716 | 16 Q7UWZ8 | Q7UWZ8 rhodospirell |
| 31 | 94 | 12.1 | 407 | 5 Q20151 | Q20151 caenorhabdi |
| 32 | 93.5 | 12.0 | 145 | 16 Q8U6N9 | Q8U6N9 agrobacteri |
| 33 | 93.5 | 12.0 | 645 | 16 Q7U1C5 | Q7U1C5 mycobacteri |
| 34 | 93.5 | 12.0 | 646 | 16 Q53818 | Q53818 mycobacteri |
| 35 | 93.5 | 12.0 | 1286 | 2 Q841Y5 | Q841Y5 campylobact |
| 36 | 93.5 | 12.0 | 1333 | 16 Q8PD38 | Q8PD38 xanthomonas |
| 37 | 92.5 | 11.9 | 151 | 16 Q7UC21 | Q7UC21 shigella fl |
| 38 | 92.5 | 11.9 | 154 | 16 Q89J15 | Q89J15 bradyrhizob |
| 39 | 92.5 | 11.9 | 160 | 16 Q8CW64 | Q8CW64 escherichia |
| 40 | 92.5 | 11.9 | 160 | 16 Q83RU7 | Q83RU7 shigella fl |
| 41 | 92.5 | 11.9 | 453 | 5 Q9NGF7 | Q9NGF7 drosophila |
| 42 | 92.5 | 11.9 | 453 | 5 Q9N6M8 | Q9N6M8 drosophila |
| 43 | 92 | 11.9 | 196 | 10 Q22638 | Q22638 zea mays (m |
| 44 | 91.5 | 11.8 | 151 | 2 Q7X238 | Q7X238 enterobacte |
| 45 | 91.5 | 11.8 | 453 | 5 Q9NGF6 | Q9NGF6 drosophila |

ALIGNMENTS

RESULT 1

| | | | | |
|--------|---|---|--------------------|-------------------|
| O33802 | O33802 | PRELIMINARY; | PRT; | 152 AA. |
| ID | O33802 | | | |
| AC | O33802; | | | |
| DT | 01-JAN-1998 | (T-EMBLrel. 05, Created) | | |
| DT | 01-JAN-1998 | (T-EMBLrel. 05, Last sequence update) | | |
| DT | 01-DEC-2001 | (T-EMBLrel. 19, Last annotation update) | | |
| DE | Agfa protein (Fragment). | | | |
| GN | AGFA. | | | |
| OS | Salmonella typhimurium. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | | |
| OC | Enterobacteriaceae; Salmonella. | | | |
| OX | NCBI_TaxID=602; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=98053981; PubMed=9393832; | | | |
| RA | Sukopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D., | | | |
| RA | Normark S.J., Rhen M.; | | | |
| RT | "Expression of thin, aggregative fimbriae promotes interaction of | | | |
| RT | Salmonella typhimurium SR-11 with mouse small intestinal epithelial | | | |
| RT | cells."; | | | |
| RL | Infect. Immun. 65:5320-5325(1997). | | | |
| DR | EMBL; AJ000514; CA04151.1; -. | | | |
| FT | NON TER 152 | | | |
| SQ | SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64; | | | |
| | Query Match | 86.7%; | Score 673; | DB 2; Length 152; |
| | Best Local Similarity | 89.4%; | Pred. No. 3.7e-47; | |
| | Matches 135; | Conservative 3; | Mismatches 13; | Indels 0; Gaps 0; |
| QY | 1 | MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSSGPDSTLSIYQGSANAALALQ 60 | | |
| Db | 1 | MKLLKVAFAAIVVSGSAVAGVVPQWGGGNNHNGSSGPDSTLSIYQGSANAALALQ 60 | | |
| QY | 61 | SDARKSETTITQSGYNGADVGQADNYDQIVTVVTHEMAHADQWNAKSDITVGYGG 120 | | |
| Db | 61 | SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQWNAKSDITVGYGG 120 | | |
| QY | 121 | NNAALVNOTADSSVMVRQVGFNNATANQY 151 | | |

```

Db      121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
RESULT 2
Q7X243 ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 73.0%; Score 566.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 1.6e-38;
Matches 117; Conservative 11; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGNGHGGSGNSGPDSTLSIYQYGSANAALQ 60
|||||
Db 1 MKLLQVAFAAIVVSGSALAGVVPQWGGGGG-GGGGSSGSPSTLSIYQYGVNNAALQ 59
|||||
QY 61 SDARKSETTITQSGYNGADYQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
|||||
Db 60 SDARKSDTHIHQGGNGADYQGGSDNSTIDITQFGKNNATIDQWNGKNSDITVQYGG 119
|||||

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLMVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240 ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.3%; Score 538; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 3.2e-36;
Matches 109; Conservative 17; Mismatches 23; Indels 2; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSGNSGPDSTLSIYQYGSANAALQ 60
|||||

```

```

Db      1 MKLLKVAFAAIVVSGSALAGVVPQW-GGNHGGSGNSGPDSSLSIYQYGSNNAALQ 58
61 SDARKSETTITQSGYNGADYQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
|||||
Db 59 SDARKSDVTITQHGNGGNGAVVVGQGADDSITSLKQTGFQNSATIDQWNAKNSDITVQYGG 118
|||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 119 NGALVNQTASDSNVLIQQVQVGFNNATANQY 149
|||||

RESULT 4
Q8CW63 ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli";
PL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 64.0%; Score 496.5; DB 16; Length 152;
Best Local Similarity 67.8%; Pred. No. 7.7e-33;
Matches 103; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGNGHGGSGNSGPDSTLSIYQYGSANAALQ 59
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGHGGSGNSGPNSELNIYQYGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADYQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 119
|||||
Db 61 QADARNSDLTITQHGNGGNGADYQGGSDSSIDLTORGFGNSATLDQWNGKDSMTVTKQFG 120
|||||
QY 120 GNNALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 121 GGNGAVIDQTASNSSVNVTVQVGFNNATANQY 152
|||||

RESULT 5
Q7X237 ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 508BB2D872DF15F3 CRC64;

Query Match 54.3%; Score 421.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 9.2e-27;
Matches 89; Conservative 25; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVFPQGGGNNHGGSSGPDSTLSIYQYGSANAALALQ 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKFIKVAALAAIVVSGSAAVAGMIQ-CGGHCHGCHGGVGGPSTLNIYQGGNSALALQ 59
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 SDARKSETTITQSGYNGADYVQGDADNYDQIVTRVTHEMAHADQWNAKSDIIVGYGG 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 TDARNSVLNISTQGGNGADYVQGGSDSSINLTQNGFNSATLDQWNSKDSVMNVSYGG 119

QY 121 NNAALVNQATDSSVMVQVFGNNATANQY 151
Db 120 LNGALVDQATSNISVTVTQIGFNGHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RL colony morphology and expression of SEF17 fimbriae.";
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1
FT NON TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 39.4%; Score 305.5; DB 2; Length 76;
Best Local Similarity 54.5%; Pred. No. 1e-17;
Matches 66; Conservative 4; Mismatches 6; Indels 45; Gaps 1;

QY 30 GNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYVQGDADNYD 89
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GNHXGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYVQGDADN-- 58

QY 90 QIVTRVTHEMAHADQWNAKSDIIVGYGNNALVNQATDSSVMVQVFGNNATAN 149
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 59 -----STIELTONGFRNATID 75

QY 150 Q 150
Db 76 Q 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).

```

```

GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curlin of Escherichia coli O78:K80 isolates associated with
RT IS1 inserti on in csbg and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29
FT NON TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0025;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVFPQGGG 29
   |||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAALAAIVVSGSALAGVFPQGGG 29

RESULT 8
Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; S00866; -.
RW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.8%; Score 115; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.061;
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 2;

QY 39 SGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYVQGDADNYDQIVTRVTH 98
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 41 SGRDNLIDLVQQTANGGI-----VFQSGSDNSAYVTQAGNDNISLVTOIGT- 87

QY 99 EMAHADQWNAKSDIIVGYGNNALVNQATDSSVMVQVFGN 144
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 88 -----NNEVQLLVQGAQNKASITQIGNDNLVQLNQLSGN 122

RESULT 9

```

```

Q89J14
ID Q89J14 PRELIMINARY; PRT; 130 AA.
AC Q89J14;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BIL5299 protein.
GN BIL5299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
CX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Iidesawa K., Iriiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.2%; Score 110; DB 16; Length 130;
Best Local Similarity 25.8%; Pred. No. 0.14;
Matches 39; Conservative 28; Mismatches 56; Indels 28; Gaps 5;

QY 4 LKVAFAAIVVSGSALAGVVPQGGNGHNGSGSPDSTLSIYQVGSANAALALQSDA 63
Db 1 MRITVLVATAIALSALTVDAA-----AGNSA-----SVLPQGITNSSFISQIGS 45

QY 64 RKSETITQSGYNGADVGQGGADNYQLVTRVVTHEMAHADQW-----NAKNSDITVQYG 119
Db 46 TSNNAITLQFGATNTATTTLQTS-----LLTVNTAVTGGGTATATNTALT-GQVG 96

QY 120 GNAALVNQTSDDSSVMVQVQFGNNATANO 150
Db 97 GNSSLIGQIGANTAGVQLGLNGSTILQ 127

RESULT 10
Q8EIH4
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
CX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imbrail M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neallson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53941.1; -.

DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 14.1%; Score 109.5; DB 16; Length 502;
Best Local Similarity 23.8%; Pred. No. 0.79; 50; Indels 49; Gaps 5;
Matches 39; Conservative 26; Mismatches 26;

QY 29 GGNHNGGSGSPDSTLSIYQVGSANAA-----LALQSDARKSETTITQSGYNG 78
Db 347 GDNNELVAFATGEDNSIEISQEGDANFAYVDATGNDNEVNVQDGDQNETIIVG-NNN 405

QY 79 ADVG-----QCADNYQLVTRVVTHEMAHADQW-----N 107
Db 406 ADVTALQHRGDLNLILIEGDENAAEIT-----QAGSGNVVGGSDSSFAASSFGVS 458

QY 108 AKNSDITVQYGGGNAALVNQTSDDSSVMVQVQFGNNATANOY 151
Db 459 GDNNSLMITQGTNDNLVLGSAQGNNSISVTSQSDMNVATVQY 502

RESULT 11
Q94821
ID Q94821 PRELIMINARY; PRT; 1748 AA.
AC Q94821; P92145; P92144; P92143; P92142; P92141; Q94820;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CNJB protein.
GN CNJB.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
CX NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88189811; PubMed=3357771;
RA Martindale D.W., Taylor F.M.;
RT "Multiple introns in a conjugation-specific gene from Tetrahymena
RT thermophila."
RL Nucleic Acids Res. 16:2189-2201(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94051569; PubMed=8233798;
RA Taylor F.M., Martindale D.W.;
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein
RT encoded by cnjB, a Tetrahymena gene active during meiosis."
RL Nucleic Acids Res. 21:4610-4614(1993).
DR EMBL; X06462; CAB37323.1; -.
DR PIR; S42136; S42136.
DR HSSP; P05888; IAAF.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf_CCHC; 7.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 7.
DR PROSITE; P850158; ZF_CCHC; 7.
DR CONFLICT 251 251 M -> I (IN REF. 1).
DR CONFLICT 256 256 I -> N (IN REF. 1).
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 14.0%; Score 109; DB 5; Length 1748;
Best Local Similarity 33.6%; Pred. No. 3.9;
Matches 37; Conservative 11; Mismatches 32; Indels 30; Gaps 6;

QY 25 QWGGGNGHNGG---GNSSGPDSTLSIYQVGSANAALALQSDARKSETTIT---QSGYGN 77
Db 1640 QFGGGNGSGGSGWTSSGSDWN-----QGSNVQESTTSSGGSGSGSGN 1685

QY 78 GADVGGGA-DNYDQLVTRVVTHEMAHADQWNAKNSDITVQG---YGNNA 124
Db 1686 QTGGGWSNDNQOQ-----QNETGGGGGSGNSQNTNNESSWGSNNQA 1729

```



```

RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646061; CAD14589.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; Fil haemagg.
DR InterPro; IPR008638; Haemagg act.
DR Pfam; PF05594; Fil haemagg; 20.
DR Pfam; PF05860; Haemagg act; 1.
DR PROSITE; PS01070; NUCLEASE_NON_SPSC; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290841C599018A107 CRC64;
SQ
Query Match 13.7%; Score 106.5; DB 16; Length 3501;
Best Local Similarity 29.5%; Pred. No. 14;
Matches 43; Conservative 19; Mismatches 45; Indels 39; Gaps 7;
QY 15 SCSALAGVVPQMGGGNNGG-NSGGPDTLSIYQGSANAA-----56
DB 2431 SCSHFTAGFSWGLGRNVGGPNSSG---VGLAPYGSAAHNAAGNSSQNASVWIG 2486
QY 57 LALOSDARKSETTITQSGYNGADVG---QCADNVQDLVTRVVTHEMA--HADQWNAKN 110
DB 2487 KSVQVQARYGDTIVSGSGISALSDVDLLAKQK-----VDIVAGNDTSRSHD---H 2535
QY 111 SDITVQYGGNNAALVNQTASDSSVM 136
DB 2536 SDRTIGDLGGYSGTVGVSASSTL 2561
RESULT 14
Q8XSD6 PRELIMINARY; PRT; 3552 AA.
ID Q8XSD6
AC Q8XSD6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSP0540.0R RS06117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OC NCBI_TaxID=305;
CX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646079; CAD17691.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; Fil haemagg.
DR InterPro; IPR008638; Haemagg act.
DR Pfam; PF05594; Fil haemagg; 20.
DR Pfam; PF05860; Haemagg act; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290841C599018A107 CRC64;
SQ
Query Match 14.0%; Score 109; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 40; Conservative 20; Mismatches 52; Indels 36; Gaps 7;
QY 11 AIVVSGSALAGVVPQ--WGGGNNHNGGNSGPDSTLSIYQGS-----ANAA----- 56
DB 693 AIAATAGAGVGLAQSIGGGGN---GGNATGGDAGFGSFQIGGGGGGGVANTANVGFK 749
QY 57 -LALOSDARKSETTITQSGYNGADVGQADNVQDLVTRVVTHEMAHADQWNAKNSDITV 115
DB 750 GLTLTQGSAAIGVIAQS-VGGGGTGTGTSYSAGI-----GFTASVAV 793
QY 116 GQYGGNNA--LVNQTASDSSVWVRQV 141
DB 794 GGTGGNGAGGVEVSVLDSAIRTCGG 821
RESULT 13
Q8Y106 PRELIMINARY; PRT; 3501 AA.
ID Q8Y106
AC Q8Y106;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RS06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OC NCBI_TaxID=305;
CX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Fuenler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti.";
proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -.
DR PIR; E95985; E95965.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004501; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
DR Plasmid; Hypothetical protein; Complete proteome.
KW SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;
SQ
Query Match 14.0%; Score 109; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 40; Conservative 20; Mismatches 52; Indels 36; Gaps 7;
QY 11 AIVVSGSALAGVVPQ--WGGGNNHNGGNSGPDSTLSIYQGS-----ANAA----- 56
DB 693 AIAATAGAGVGLAQSIGGGGN---GGNATGGDAGFGSFQIGGGGGGGVANTANVGFK 749
QY 57 -LALOSDARKSETTITQSGYNGADVGQADNVQDLVTRVVTHEMAHADQWNAKNSDITV 115
DB 750 GLTLTQGSAAIGVIAQS-VGGGGTGTGTSYSAGI-----GFTASVAV 793
QY 116 GQYGGNNA--LVNQTASDSSVWVRQV 141
DB 794 GGTGGNGAGGVEVSVLDSAIRTCGG 821
RESULT 13
Q8Y106 PRELIMINARY; PRT; 3501 AA.
ID Q8Y106
AC Q8Y106;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RS06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OC NCBI_TaxID=305;
CX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Fuenler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti.";
proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -.
DR PIR; E95985; E95965.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004501; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
DR Plasmid; Hypothetical protein; Complete proteome.
KW SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;
SQ
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-24
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : 1_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 775 | 100.0 | 151 | 3 | AAB36352 |
| 2 | 714 | 92.1 | 151 | 3 | AAB36347 |
| 3 | 700 | 90.3 | 151 | 3 | AAB36355 |
| 4 | 690 | 89.0 | 151 | 2 | AAR74625 |
| 5 | 690 | 89.0 | 151 | 3 | AAB36341 |
| 6 | 689 | 88.9 | 151 | 2 | AAW23570 |
| 7 | 655 | 84.5 | 151 | 3 | AAB36346 |
| 8 | 609 | 78.6 | 151 | 3 | AAB36353 |
| 9 | 609 | 78.6 | 151 | 3 | AAB36351 |
| 10 | 608 | 78.5 | 151 | 3 | AAB36349 |
| 11 | 606 | 78.2 | 151 | 3 | AAB36350 |
| 12 | 600 | 77.4 | 151 | 3 | AAB36354 |
| 13 | 575 | 74.2 | 151 | 3 | AAB36348 |
| 14 | 533 | 68.8 | 151 | 3 | AAB36343 |
| 15 | 528 | 68.1 | 151 | 7 | ABR2651 |
| 16 | 505 | 65.2 | 120 | 2 | AAW23569 |
| 17 | 505 | 65.2 | 120 | 2 | AAW23569 |
| 18 | 455 | 58.7 | 142 | 2 | AAW23564 |
| 19 | 383 | 49.4 | 122 | 2 | AAW23563 |
| 20 | 237 | 30.6 | 45 | 3 | AAW23561 |
| 21 | 132 | 17.0 | 22 | 3 | AAW23561 |
| 22 | 123 | 15.9 | 23 | 3 | AAW23561 |
| 23 | 123 | 15.9 | 23 | 3 | AAW23561 |
| 24 | 123 | 15.9 | 23 | 3 | AAW23561 |
| 25 | 115 | 14.8 | 22 | 3 | AAW23561 |

| | | | | | |
|----|------|------|------|---|----------|
| 26 | 115 | 14.8 | 22 | 3 | AAB36339 |
| 27 | 115 | 14.8 | 22 | 3 | AAB36320 |
| 28 | 113 | 14.6 | 24 | 7 | ABR2644 |
| 29 | 109 | 14.1 | 23 | 3 | AAB36340 |
| 30 | 109 | 14.1 | 23 | 3 | AAB36324 |
| 31 | 109 | 14.1 | 23 | 3 | AAB36319 |
| 32 | 98.5 | 12.7 | 151 | 3 | AAB36342 |
| 33 | 98 | 12.6 | 26 | 7 | ABR2649 |
| 34 | 98 | 12.6 | 26 | 7 | ABR2645 |
| 35 | 98 | 12.6 | 903 | 2 | AAW35006 |
| 36 | 96.5 | 12.5 | 151 | 3 | AAB36344 |
| 37 | 96 | 12.4 | 19 | 3 | AAB36323 |
| 38 | 96 | 12.4 | 19 | 3 | AAB36336 |
| 39 | 96 | 12.4 | 19 | 3 | AAB36328 |
| 40 | 92.5 | 11.9 | 850 | 4 | ABR2649 |
| 41 | 92.5 | 11.9 | 1028 | 4 | ABR2648 |
| 42 | 92 | 11.9 | 23 | 3 | AAB36331 |
| 43 | 91 | 11.7 | 688 | 5 | ABP74039 |
| 44 | 90.5 | 11.7 | 378 | 4 | ABR2646 |
| 45 | 90.5 | 11.7 | 520 | 6 | AAO16497 |

ALIGNMENTS

RESULT 1
AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200006102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PY (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX
PM 2000-672631/65.
XX
PS N-PSDB; AAC4628.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation dependent assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.6e-68;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNYDQLVTRVVT 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNYDQLVTRVVT 120
 QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151

RESULT 2
 AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#2 amino acid sequence SEQ ID NO:14.
 KW Salmonella; agfA; chromosomal gene replacement; fimbtrin; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI: 2000-672631/65.
 XX N-PSDB; AAC64623.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 XX which encodes foreign epitope or antigen, expresses recombinant AgfA
 XX protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsaA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 1.8e-62;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNYDQ 113
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNYDQ 120
 QY 114 LVTRVVTHEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 LVTRVVTHEMAHA-----SVNVRQVGFNNATANQY 151

RESULT 3
 AAB36355
 ID AAB36355 standard; protein; 151 AA.
 XX AAB36355;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#10 amino acid sequence SEQ ID NO:30.
 KW Salmonella; agfA; chromosomal gene replacement; fimbtrin; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.

```

XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64631.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant AgfA
XX PT protein useful for eliciting immune response in animal.
XX PS Disclosure; Page 139; 139pp; English.
XX CC The present invention describes a recombinant agfA gene (I) where a
XX CC segment of the gene has been replaced by a segment of a foreign DNA
XX CC sequence which encodes a foreign epitope or antigen. Also described are:
XX CC (1) use of thin aggregative fimbriae (SfH17/TAf) nucleation depended
XX CC assembly system of strains of Salmonella, Escherichia coli and
XX CC Enterobacteriaceae for the production of fimbriae comprising recombinant
XX CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX CC directing recombination of a recombinant gene into the chromosome of the
XX CC homologous species; (3) directing recombination of a recombinant gene
XX CC back into the chromosome of the homologous species, replacing the native
XX CC copy of that gene; and (4) eliciting an immune response in an animal,
XX CC comprising separating an amino acid polymer comprising a recombinant AgfA
XX CC protein containing a replacement segment or segments of foreign amino
XX CC acid sequence or sequences grown on a Salmonella, E. coli or
XX CC Enterobacteriaceae host cell, from the host cell and introducing the
XX CC polymer into the animal in conjunction with a carrier or diluent. (I) is
XX CC useful for the expression of recombinant AgfA protein which is useful for
XX CC eliciting an immune response in an animal. In a fimbrial presentation
XX CC system the heterologous antigens are presented in high numbers (up to
XX CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX CC immunogenicity and adhesion properties relevant for an efficient live
XX CC vaccine, the carrier fimbrial subunit proteins are usually strong
XX CC immunogens, which may be important for directing an immune response
XX CC against the inserted epitope, and hybrid fimbriae are easy and
XX CC inexpensive to purify in large amount. The present sequence is given in
XX CC the exemplification of the present invention
XX SQ Sequence 151 AA;
    Query Match 90.3%; Score 700; DB 3; Length 151;
    Best Local Similarity 89.9%; Pred. No. 4.2e-61;
    Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;
    QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
    DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
    QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNYDQLVTRVVT 120
    DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNYDQLVTRVVT 120
    QY 121 HEMAHAA-----NCTASDSSVMVROVGFGNNTATNOY 151
    DB 113 HEMAHAGGNAALVNQTASDSSVMVROVGFGNNTATNOY 151
    RESULT 4
    AAR74625
    ID AAR74625 standard; protein; 151 AA.
    XX AAR74625;
    XX 25-MAR-2003 (revised)
    DT 26-JUN-1995 (first entry)
    XX AgfA sequence.
    XX Salmonella; AgfA; vaccine.
    XX Salmonella.
    OS
    DR
    PN WO94255598-A2.
    XX 10-NOV-1994.
    XX 26-APR-1994; 94WO-IB000207.
    XX 26-APR-1993; 93US-00054452.
    XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
    PA (KING/) KING J.
    XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
    XX WPI; 1994-358275/44.
    XX N-PSDB; AAQ87467.
    XX Eliciting an immune response to Salmonella - using attenuated Salmonella
    PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
    PT Disclosure; Fig 7B; 95pp; English.
    XX The Salmonella AgfA protein and DNA are used in vaccine and genetic
    CC immunization compositions, respectively, to elicit an immune response to
    CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
    CC on 25-MAR-2003 to correct PN field.)
    XX Sequence 151 AA;
    Query Match 89.0%; Score 690; DB 2; Length 151;
    Best Local Similarity 91.4%; Pred. No. 4.1e-60;
    Matches 139; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
    QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
    DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
    QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNYDQLVTRVVT 120
    DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNSDITVQYGG 120
    QY 121 HEMAHANQTASDSSVMVROVGFGNNTATNOY 151
    DB 121 NNAALVNQTASDSSVMVROVGFGNNTATNOY 151
    RESULT 5
    AAB36341
    ID AAB36341 standard; protein; 151 AA.
    XX AAB36341;
    XX 26-FEB-2001 (first entry)
    DT
    DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
    KW Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
    KW vaccine; immune response; immunogen.
    XX Salmonella enteritidis.
    OS
    PN WO2000060102-A2.
    XX 12-OCT-2000.
    XX 05-APR-2000; 2000WO-CA000356.
    XX 05-APR-1999; 99US-0127888P.
    XX (UYVI-) UNIV VICTORIA.
    PA White AP, Doran JL, Collison SK, Kay WW;
    PI WPI; 2000-672631/65.
    DR

```

DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 89.0%; Score 690; DB 3; Length 151;
 Best Local Similarity 91.4%; Pred. No. 4.1e-60;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGGADNSTIELTQGFRRNATIDQWNAKNDQLVTRVVT 120
 DB 61 SPARKSETTITQSGYNGADVGGADNSTIELTQGFRRNATIDQWNAKNSDITVQYGG 120
 QY 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 6
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 XX
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 XX

PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;
 XX
 DR WPI; 1997-309886/28.
 XX
 DR N-PSDB; AAT74142.
 XX
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 PS Example 2; Fig 7; 85pp; English.
 CC
 CC The present sequence represents agfa encoded by the full agfa gene
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
 CC bacteria of the family Enterobacteria. It can also be used to provide
 CC proteins and antibodies which can be used for assays. The nucleic acid
 CC sequence can be used to provide probes or primers which can specifically
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
 CC strains that are pathogenic to warm-blooded animals relative to nucleic
 CC acid molecules from virtually all other microbial organisms. (Updated on
 CC 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 151 AA;

Query Match 88.9%; Score 689; DB 2; Length 151;
 Best Local Similarity 91.4%; Pred. No. 5.2e-60;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGGADNSTIELTQGFRRNATIDQWNAKNDQLVTRVVT 120
 DB 61 SPARKSETTITQSGYNGADVGGADNSTIELTQGFRRNATIDQWNAKNSDITVQYGG 120
 QY 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNPALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7
 AAB36346
 ID AAB36346 standard; protein; 151 AA.
 XX
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbriin; epitope;
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX

```
PR 05-APR-1999; 99US-0127888P.
XX (UVVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 135; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 84.5%; Score 655; DB 3; Length 151;
XX Best Local Similarity 79.8%; Pred. No. 1.2e-56;
XX Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
XX
XX QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDOWNAK----- 109
XX DB 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
XX
XX QY 110 -----NYDOLVTRVTHEMAHANQTASDSSVMVQVGFNNATANQY 151
XX DB 121 NNAALVNYDQVTRVTHEMAHA-----NNATANQY 151
XX
XX RESULT 8
XX AAB36353 standard; protein; 151 AA.
XX AC AAB36353;
XX DT 26-FEB-2001 (first entry)
XX DE AgfA::PT3#8 amino acid sequence SEQ ID NO:26.
XX XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX KW vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA0000356.
XX 05-APR-1999; 99US-0127888P.
XX (UVVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 78.6%; Score 609; DB 3; Length 151;
XX Best Local Similarity 82.1%; Pred. No. 4.2e-52;
XX Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNYDQVTRVVT 120
XX DB 61 LVTRVTHEMAHAGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
XX
XX QY 121 HEMAHANQTASDSSVMVQVGFNNATANQY 151
XX DB 121 NNAALVNYDQVTRVTHEMAHA-----NNATANQY 151
XX
XX RESULT 9
```

AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64627.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 XX
 Query Match 78.6%; Score 609; DB 3; Length 151;
 Best Local Similarity 73.6%; Pred. No. 4.2e-52;
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGGSSGPDSTLSIYQVGSANAALQ 60
 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGGSSGPDSTLSIYQVGSANAALQ 60

QY 61 SDARKSETTTTQSGYNGADVGOGADNNTIELTQNGFRNNATIDQWNAKVQDQVTRVVT 120
 DB 61 SDARKSETTTTQSGYNGADVGOGADNNTIELTQNGFRNNATIDQWNAKVQDQVTRVVT 97
 QY 121 HEMAHA-----NOTASDSVMVRQYVGFNGNATANQY 151
 DB 98 HEMAHADQWNAKNSDITVGYQVGGNNAALVNOTASDSVMVRQYVGFNGNATANQY 151

RESULT 10
 AAB36349
 ID AAB36349 standard; protein; 151 AA.
 XX
 AC AAB36349;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
 DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64625.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 78.5%; Score 608; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 5.2e-52;
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120

QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 11
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64626.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 137; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/1AF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 78.2%; Score 606; DB 3; Length 151;
 Best Local Similarity 82.1%; Pred. No. 8.2e-52;
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDITVGYGG 120

QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 12
 AAB36354
 ID AAB36354 standard; protein; 151 AA.
 XX
 AC AAB36354;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 XX
 DR N-PSDB; AAC64630.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 3.2e-51;
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNDITVGVYGG 120

Db 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNDITVGVYGG 120

QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 13

AAB36348

ID AAB36348 standard; protein; 151 AA.

AC AAB36348;

DT 26-FEB-2001 (first entry)

DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended are:
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.2%; Score 575; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 9.5e-49;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNDITVGVYGG 120

Db 61 SDARKSETTITQSGYNGADVCGADNSTIETQNGFRNNATIDQWNAKNDITVGVYGG 120

QY 121 HEMAHANQTSADSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 14

AAB36343

ID AAB36343 standard; protein; 151 AA.

AC AAB36343;

DT 26-FEB-2001 (first entry)

DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

OS Escherichia coli.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 68.8%; Score 533; DB 3; Length 151;
Best Local Similarity 70.2%; Pred. No. 1.3e-44;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGDPSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGGNGSGPNSLNLYQYGGNSALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNATIDQWNAKNDQVTRVVT 120
DB 61 TDARNSDLTITQGGNGADVGCGDDSSIDLITQRCFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
RESULT 15
ID ABR82651
XX ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX Escherichia coli.
XX WO2003064446-A2.
XX 07-AUG-2003.
XX
PF 30-JAN-2003; 2003WO-EP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
PA (HANS-) HANSA MEDICAL RES AB.
XX
PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;
XX
DR WPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
PS Disclosure; Page 41-42; 42pp; English.
XX
CC The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
SQ Sequence 151 AA;
Query Match 68.1%; Score 528; DB 7; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.2e-44;
Matches 105; Conservative 18; Mismatches 28; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGDPSTLSIYQYGSANAALQ 60
DB 1 MKLLKVEAIAAIVFSGSALAGVVPQYGGGNGGNGSGPNSLNLYQYGGNSALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNATIDQWNAKNDQVTRVVT 120
DB 61 TDARNSDLTITQGGNGADVGCGDDSSIDLITQRCFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
Search completed: August 2, 2004, 14:48:27
Job time : 44.9 secs

This Page Blank (uspto)

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYYQGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNYDQLVTRVVT 120
 Db 61 SPARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNYDQLVTRVVT 120
 QY 121 HEMAHANQASDSSVMVRQVGGNNATQY 151
 Db 121 NNPALVNTQASDSSVMVRQVGGNNATQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Cloutier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; TITLE OF INVENTION: OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

US-08-233-788A-57

Query Match 65.2%; Score 505; DB 1; Length 120;
 Best Local Similarity 88.4%; Pred. No. 1.7e-42;
 Matches 99; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGHNGGNSGPDSTLSIYYQGSANAALQSDARKSETTITQSGYNGADV 81
 Db 1 VVPQWGGGNGHNGGNSGPDSTLSIYYQGSANAALQSDARKSETTITQSGYNGADV 60
 QY 82 GQADNSTIETQNGFRNNATIDQWNAKNYDQLVTRVTHEMAHANQASDS 133
 Db 61 GQADNSTIETQNGFRNNATIDQWNAKNYDQLVTRVTHEMAHANQASDS 112

RESULT 3

US-09-196-387-8
 ; Sequence 8, Application US/09196387
 ; Patent No. 6277613
 ; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/196,387
 ; FILING DATE:
 ; CLASSIFICATION:

PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/095,225
 ; FILING DATE: June 10, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 673 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-196-387-8

Query Match 11.7%; Score 90.5; DB 3; Length 673;
 Best Local Similarity 28.4%; Pred. No. 1.1;
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYYQGSANAALQSDAR 64
 Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNKSPSSSSPTSS-SSSSSPSPGSSLAESPEAA 157
 QY 65 KSETTIT----QSGYNGADVGGADNSTIETQNGFRNNATIDQWNAKNYDQLVTRVVT 120
 Db 158 GVSSTAPLPGCAAGPGTGVPVAGSGLRELLACRNG-----DVSRRVKRLVDA--- 204
 QY 121 HEMAHANQASD-----SSVMVRQVGGF 143
 Db 205 -----ANVNAKDVAGRKSSPLHFAAGFG 227

RESULT 4

US-09-841-835-8
 ; Sequence 8, Application US/09841835
 ; Patent No. 6506587
 ; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia
 ; APPLICANT: Smith, Susan
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
 ; TITLE OF INVENTION: OF USE THEREOF
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA

Query Match 11.7%; Score 90.5; DB 3; Length 1327;
Best Local Similarity 28.4%; Pred. No. 2.8;

| | Query Match | Best Local Similarity | Score | DB | Length |
|----|-------------|---|--------|------|--------|
| | Matches | Conservative | Indels | Gaps | |
| Qy | 6 | VAAFAAI-VVSGSALAGVPWQGGGHHNGGNSGPDSTLSIVYQSANAALAQSDAR | 64 | | |
| Dd | 99 | VAAAAPPVAVTSIAAGVAAPNPAAGSGSNFSSSPSTSS-SSSSPSPGSSLAESPEAA | 157 | | |
| Qy | 65 | KSETTIT-----QSQGYMGADVGQGADNSTIETLTGNERNNATIDQMNAKYDQLVTWVT | 120 | | |
| Dd | 158 | GVASTPIGPGAAGPGTGVPASGALRELLAACRNG-----DVSRVKRLVDLA--- | 204 | | |

QY 121 HEMAHANOTASD-----SSVMVRQVGF 143
Db 205 -----ANVNKDMAGKSSPLHFAAGF 227

RESULT 9

US-09-972-115A-8
; Sequence 9, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Flatuszek A.
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match 11.7%; Score 90.5; DB 4; Length 1327;
Best Local Similarity 28.4%; Pred. No. 2.8;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;
QY 6 VAFAAI-VVGSALAGVVPQWGGGNGGNSGSPDSTLSIYQGSANAALALQSDAR 64
Db 99 VAAPVVPVAVTSSAAGVAPNPAGSGNSNPSSTSSPTSS-SSSSPSGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVQCGADNSTILTONTGFENNATIDWNKYNVDQLATRVVT 120
Db 158 GVSSTAPIGPAGAGPGTGPVAVSGALRELLACRNG-----DVSrvKRLVDA--- 204
QY 121 HEMAHANOTASD-----SSVMVRQVGF 143
Db 205 -----ANVNKDMAGKSSPLHFAAGF 227

RESULT 10

US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: 812-5 Hirano
; STREET: Ieshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 986-2340
; TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 11.5%; Score 89.5; DB 3; Length 738;
Best Local Similarity 35.4%; Pred. No. 1.6;
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;
QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYQGSANAALALQSD 62
Db 419 LLKSSASASASASASAG-----GGGGGNGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYNGADVQCG 84
Db 461 -----AALAAAGAGGLGGGG 477

RESULT 11

US-09-198-452A-171
; Sequence 171, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 171
; LENGTH: 1156
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-171

Query Match 11.2%; Score 87; DB 4; Length 1156;
Best Local Similarity 32.0%; Pred. No. 5.1;
Matches 32; Conservative 14; Mismatches 34; Indels 20; Gaps 5;
QY 44 TLSIYQGSANAALALQSDARKSETTITQSGYNGADVQCGADNSTI-----ELTQ 94
Db 859 TVSPEDYAAVQAALA--AVVRKHESLIV-STYGLGAQEGQTSSKVTTLMRDLHAVEELVE 915
QY 95 NGRFRNATIDWNKYNVDQLATRVVTHEMAHANQTASDSS 134
Db 916 MG-----VETRYLRNSDQILHRV--HSVLHSHLRDSDSS 947

RESULT 12

```
US-09-252-991A-32096
; Sequence 32096, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32096
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32096

Query Match 10.9%; Score 84.5; DB 4; Length 339;
Best Local Similarity 21.5%; Pred. No. 1.8;
Matches 34; Conservative 32; Mismatches 67; Indels 25; Gaps 5;

QY 5 KVAFAAIVVSGSALAGVVPQWGG-----GGNHGGGSGDPDSTLSIYQYGSANAL 57
DB 58 KVSNEFTL--NNASVSGSIKIDAGNVGNVAAAGDNNQQAALASADASFVGTATAS- 114
QY 58 ALQSDARKSETTITQSGYNGGADVGOGADNSTIELTQNGFRNATIDQWNAKNDQ---- 113
DB 115 -----TSVLOGYGTNTLVNPNNTSLNSANNVSGNLGVNV-ACGNFNOQKND 163
QY 114 LVTRVVTHEMAHANTASDSSVVMVQVGFNNATANQY 151
DB 164 LAAAVSNGQYSTAGSAASQSTGTNTVNSANYVGTY 201

RESULT 13
US-09-336-447A-15
; Sequence 15, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-15

Query Match 10.7%; Score 83; DB 4; Length 889;
Best Local Similarity 27.9%; Pred. No. 8.9;
Matches 39; Conservative 16; Mismatches 51; Indels 34; Gaps 8;

QY 28 GGNHGGGSGDPDSTLSIYQ---YGSANALALQSDARKSETTITQSGY-----GNCA 79
DB 69 GAGRHNNGGSAHHSILGGWNTVNGYTSALVGGYGNETQDGYTFVGGGYNKLAKNYT 128
QY 80 DVGQG-----ADNSTIELTQNGFRN-----NATID---QWNAKNDQLVTRVVTHEMAH 125
DB 129 FVGGGYKNLAEGDNATI---AGGFANLAEAGDNTIAGGFENRAEGIDSVWSG-----GY 179
```

```
QY 126 ANQTASDSSVMVRQVGFNN 145
DB 180 ANQATGESSTV---AGGSNN 196

RESULT 14
US-07-731-157A-2
; Sequence 2, Application US/07731157A
; Patent No. 5457032
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lehting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/731,157A
; FILING DATE: 19910509
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D., BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/00US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-731-157A-2

Query Match 10.5%; Score 81.5; DB 1; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9;

QY 1 MKLLKVAFAAIVVSGSALAGVV-----PQ-----WGGGNGHGGG 36
DB 2 LRVLHRAASALVMTVIGLAPAVAFALABPTSTPQAPIAAYKPRSNELWDG---YGV 57
QY 37 NSSGPDSTLSIYQYGSANA-----ALALQSDARKSETTITQSGYNGGADV-GQGADNST 89
DB 58 HIYGVDPASAFYGYGWAQARSQDNLRLYGEAR-----GKGEYWGPDYEQT 106
QY 90 IELTQNGFRNATIDQWNAK-----NYDQLVTRVVTHEMAHANTASDSSVMVQV 140
DB 107 VLLLTNGVPERA--QQWYAAQQSPDFRANLDAFAAGI---NAYAAQNPDDISPDVRQV 158

RESULT 15
US-08-541-780-2
; Sequence 2, Application US/08541780
; Patent No. 5935831
; GENERAL INFORMATION:
```

APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GBRO-027/000US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 720 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-541-780-2

Query Match 10.5%; Score 81.5; DB 2; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9;
QY 1 MKLLKVAFAAIVVSGSALAGV-----PQ-----WCGGNGHNGG 36
Db 2 LRVLRASALVMTVIGLAPAVAFALAEPTSTQAPTAAYKPRSNEILMDG----YGV 57
QY 37 NSSGPDSTLSIYQGSANA-----ALALQSDARKSETTITQSGYNGADV-GQCADNST 89
Db 58 HIYGVDAFSAFYGYGWAQARSQGDNIILRYGEAR-----GKGAEWGPDYEQT 106
QY 90 IELTONGFRNATIDWNK-----NYDOLVTRVTHENAHAHQASSSVMVROV 140
Db 107 VLLTNGVPERA--QQWYAQQSPDFRANLDAFAAGI-----NAYAQQNPDISPDVRQV 158

Search completed: August 2, 2004, 14:58:35
Job time : 12 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:**
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:**
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap:**
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap:**
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:**
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap:**
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap:**
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap:**
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:**
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap:**
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap:**
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap:**
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap:**
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap:**
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap:**
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap:**
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap:**
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap:**
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 530 | 68.4 | 151 | 12 | US-09-741-873B-4 |
| 2 | 530 | 68.4 | 151 | 12 | US-09-741-873B-4 |
| 3 | 452 | 58.3 | 131 | 12 | US-09-741-873B-2 |
| 4 | 452 | 58.3 | 131 | 12 | US-09-741-873B-2 |
| 5 | 111 | 14.3 | 445 | 15 | US-10-369-493-20638 |
| 6 | 91 | 11.7 | 688 | 14 | US-10-032-585-7876 |
| 7 | 90.5 | 11.7 | 263 | 12 | US-10-425-114-49960 |
| 8 | 90.5 | 11.7 | 673 | 9 | US-09-841-835-8 |
| 9 | 90.5 | 11.7 | 949 | 9 | US-09-841-835-10 |
| 10 | 90.5 | 11.7 | 1327 | 9 | US-09-841-835-2 |
| 11 | 90.5 | 11.7 | 1327 | 10 | US-09-972-115A-8 |
| 12 | 90.5 | 11.7 | 1327 | 14 | US-10-159-937-4 |
| 13 | 89 | 11.5 | 369 | 12 | US-10-425-114-56041 |
| 14 | 89 | 11.5 | 486 | 12 | US-10-424-599-275468 |
| 15 | 89 | 11.5 | 507 | 12 | US-10-425-114-57763 |

| | | | | | | |
|----|------|------|------|----|----------------------|-------------------|
| 16 | 88.5 | 11.4 | 145 | 16 | US-10-437-963-147748 | Sequence 147748, |
| 17 | 88.5 | 11.4 | 438 | 14 | US-10-156-761-9343 | Sequence 9343, Ap |
| 18 | 87 | 11.2 | 1156 | 15 | US-10-289-762-171 | Sequence 171, App |
| 19 | 86 | 11.1 | 186 | 12 | US-10-282-122A-49412 | Sequence 49412, A |
| 20 | 86 | 11.1 | 482 | 14 | US-10-156-761-8763 | Sequence 8763, Ap |
| 21 | 86 | 11.1 | 762 | 12 | US-10-282-122A-59874 | Sequence 59874, A |
| 22 | 85.5 | 11.0 | 270 | 16 | US-10-437-963-122263 | Sequence 122263, |
| 23 | 85.5 | 11.0 | 271 | 14 | US-10-156-761-11721 | Sequence 11721, A |
| 24 | 85.5 | 11.0 | 2457 | 12 | US-10-282-122A-49854 | Sequence 49854, A |
| 25 | 85 | 11.0 | 478 | 16 | US-10-437-963-115033 | Sequence 115033, |
| 26 | 84 | 10.8 | 254 | 10 | US-09-880-748-1136 | Sequence 1136, Ap |
| 27 | 84 | 10.8 | 254 | 10 | US-09-880-748-1136 | Sequence 1136, Ap |
| 28 | 84 | 10.8 | 254 | 12 | US-10-293-418-1136 | Sequence 1136, Ap |
| 29 | 84 | 10.8 | 254 | 12 | US-10-293-418-1136 | Sequence 1136, Ap |
| 30 | 83.5 | 10.8 | 350 | 14 | US-10-156-761-13168 | Sequence 13168, A |
| 31 | 83.5 | 10.8 | 394 | 16 | US-10-437-963-163379 | Sequence 163379, |
| 32 | 83 | 10.7 | 242 | 12 | US-10-425-114-61520 | Sequence 61520, A |
| 33 | 83 | 10.7 | 889 | 10 | US-09-952-267-15 | Sequence 15, Appl |
| 34 | 82.5 | 10.6 | 1074 | 12 | US-10-282-122A-50616 | Sequence 50616, A |
| 35 | 82 | 10.6 | 214 | 16 | US-10-437-963-119143 | Sequence 119143, |
| 36 | 82 | 10.6 | 254 | 10 | US-09-880-748-1226 | Sequence 1226, Ap |
| 37 | 82 | 10.6 | 254 | 12 | US-10-293-418-1226 | Sequence 1226, Ap |
| 38 | 82 | 10.6 | 678 | 12 | US-10-282-122A-54573 | Sequence 54573, A |
| 39 | 82 | 10.6 | 883 | 12 | US-10-282-122A-42626 | Sequence 42626, A |
| 40 | 81.5 | 10.5 | 189 | 16 | US-10-437-963-170736 | Sequence 170736, |
| 41 | 81.5 | 10.5 | 256 | 14 | US-10-156-761-11972 | Sequence 11972, A |
| 42 | 81.5 | 10.5 | 429 | 16 | US-10-437-963-124678 | Sequence 124678, |
| 43 | 81.5 | 10.5 | 449 | 12 | US-10-424-599-285485 | Sequence 285485, |
| 44 | 81.5 | 10.5 | 509 | 12 | US-10-425-114-49393 | Sequence 48393, A |
| 45 | 81.5 | 10.5 | 529 | 12 | US-10-425-114-49406 | Sequence 49406, A |

ALIGNMENTS

RESULT 1

US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 68.4%; Score 530; DB 12; Length 151;

Best Local Similarity 69.8%; Pred. No. 7.2e-46;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QV 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDTLSIYQYGSANAALALQ 60

```
Db 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGGNGSPNSLNIIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQWNAKNDYDQVTRVVT 120
Db 61 TDARNSDLTITQGGGNGADVGGSDDSIDLTQRFNGNSATLDQWNGKNSMTVKQFGG 120
QY 121 HEMAHANQASDSSVMVRQVGFNNATANOY 151
Db 121 GNCAAVDQTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 2

```
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
```

```
US-09-741-873B-2
Query Match 58.3%; Score 452; DB 12; Length 131;
Best Local Similarity 66.4%; Pred. No. 5e-38;
Matches 87; Conservative 18; Mismatches 26; Indels 0; Gaps 0;
```

```
QY 21 GVVPQGGGNGHNGGNGSGPDSITLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQYGGGNGHGGGNGSGPNSLNIIYQYGGNSALALQTDARNSDLTITQGGGNGAD 60
QY 81 VGGADNSTIELTQNGFRNNATIDQWNAKNDYDQVTRVTHEMAHANQASDSSVMVRQV 140
Db 61 VGGSDSDSIDLTQRFNGNSATLDQWNGKNSMTVKQFGGNGAAYDQTASNSVNVTVQ 120
QY 141 GFGNNATANOY 151
Db 121 GFGNNATAHQY 131
```

RESULT 4

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
```

```
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
```

```
US-09-741-873B-4
Query Match 68.4%; Score 530; DB 12; Length 151;
Best Local Similarity 69.5%; Pred. No. 7.2e-46;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;
```

```
QY 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGGNGSGPDSITLSIYQYGSANAALALQ 60
Db 1 MKLLKVAALAAIVFSGSAGVVPQYGGGNGGNGSPNSLNIIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQWNAKNDYDQVTRVVT 120
Db 61 TDARNSDLTITQGGGNGADVGGSDDSIDLTQRFNGNSATLDQWNGKNSMTVKQFGG 120
QY 121 HEMAHANQASDSSVMVRQVGFNNATANOY 151
Db 121 GNCAAVDQTASNSVNVTVQVGFNNATAHQY 151
```

RESULT 3

```
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
```

```
Query Match 58.3%; Score 452; DB 12; Length 131;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-24
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSVVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues
Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main: *
1: /cgn2_6/ptodata/2/paa/ptorus COMB.pcp: *
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp: *
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp: *
4: /cgn2_6/ptodata/2/paa/US080 COMB.pcp: *
5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp: *
6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp: *
7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp: *
8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp: *
9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp: *
10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp: *
11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp: *
12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp: *
13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp: *
14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp: *
15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp: *
16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp: *
17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp: *
18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp: *
19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp: *
20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp: *
21: /cgn2_6/ptodata/2/paa/US097 COMB.pcp: *
22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp: *
23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp: *
24: /cgn2_6/ptodata/2/paa/US099 COMB.pcp: *
25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp: *
26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp: *
27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp: *
28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp: *
29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp: *
30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp: *
31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp: *
32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp: *
33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | Query | Score | Match | Length | ID | Description |
|--------|-------|-------|-------|--------|----|-------------|
|--------|-------|-------|-------|--------|----|-------------|

| | | | | | | |
|----|-------|-------|------|----|----------------------|-------------------|
| 1 | 775 | 100.0 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 2 | 714 | 92.1 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 3 | 700 | 90.3 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 4 | 690 | 89.0 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 5 | 689 | 88.9 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 6 | 655 | 84.5 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 7 | 609 | 78.6 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 8 | 609 | 78.6 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 9 | 608 | 78.5 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 10 | 606 | 78.2 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 11 | 603 | 77.8 | 151 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 12 | 600 | 77.4 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 13 | 575 | 74.2 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 533 | 68.8 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 530 | 68.4 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 530 | 68.4 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 528 | 68.1 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 528 | 68.1 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 505 | 65.2 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 475 | 61.3 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 21 | 475 | 61.3 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 475 | 61.3 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 23 | 471 | 60.8 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 24 | 452 | 58.3 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 452 | 58.3 | 131 | 21 | US-09-741-873B-2 | Sequence 2, Appl |
| 26 | 348 | 44.9 | 109 | 19 | US-09-543-407-35 | Sequence 35, Appl |
| 27 | 347 | 44.8 | 68 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 28 | 248.5 | 32.1 | 70 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 29 | 237 | 30.6 | 48 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 30 | 111 | 14.3 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 31 | 111 | 14.3 | 445 | 33 | US-60-360-039-20638 | Sequence 20638, A |
| 32 | 107 | 13.8 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 33 | 107 | 13.8 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 34 | 107 | 13.8 | 186 | 30 | US-10-417-886-5833 | Sequence 5833, Ap |
| 35 | 98.5 | 12.7 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 36 | 98 | 12.6 | 903 | 24 | US-09-914-543-44 | Sequence 44, Appl |
| 37 | 96.5 | 12.5 | 151 | 19 | US-09-543-407-8 | Sequence 8, Appl |
| 38 | 92.5 | 11.9 | 850 | 20 | US-09-614-150A-24084 | Sequence 24084, A |
| 39 | 92.5 | 11.9 | 850 | 20 | US-09-614-150A-24084 | Sequence 24084, A |
| 40 | 92.5 | 11.9 | 850 | 33 | US-60-191-637-24179 | Sequence 24179, A |
| 41 | 92.5 | 11.9 | 850 | 33 | US-60-191-681-19019 | Sequence 19019, A |
| 42 | 92.5 | 11.9 | 1028 | 20 | US-09-614-150-14916 | Sequence 14916, A |
| 43 | 92.5 | 11.9 | 1028 | 20 | US-09-614-150A-14916 | Sequence 14916, A |
| 44 | 92.5 | 11.9 | 1028 | 33 | US-60-167-217-14978 | Sequence 14978, A |
| 45 | 92.5 | 11.9 | 1028 | 33 | US-60-173-464-12194 | Sequence 12194, A |

ALIGNMENTS

RESULT 1
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-24

Query Match 100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 2.7e-74; Indels 0; Gaps 0;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 92.1%; Score 714; DB 19; Length 151;
Best Local Similarity 91.1%; Pred. No. 9.2e-68; Indels 14; Gaps 2;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 113
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
QY 114 LVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 LVTRVVTHEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 90.3%; Score 700; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 2.9e-66; Indels 16; Gaps 2;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 112
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 113 HEMAHANQATSDSSVMVRQVGFNNATANQY 151

RESULT 4

US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 89.0%; Score 690; DB 19; Length 151;
Best Local Similarity 91.4%; Pred. No. 3.4e-65; Indels 0; Gaps 0;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWAKNYDQLVTRVVT 120
QY 121 HEMAHANQATSDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 5

US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.

APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
BASED VACCINES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 88.9%; Score 689; DB 6; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.4e-65;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNDLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNDLVTRVVT 120
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNPALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-12
Sequence 12, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 151

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match 84.5%; Score 655; DB 19; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.9e-61;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNDLVTRVVT 109
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNDLVTRVVT 120
QY 110 -----NYDQLVTRVVTHEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151

RESULT 7
US-09-543-407-22
Sequence 22, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 22
LENGTH: 151
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 73.8%; Pred. No. 1.6e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAITDQWNAKNDLVTRVVT 120
DB 61 SDARKSETTITQSGYNGADVGQAD-----NYDQLVTRVVT 97
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 98 HEMAHANQASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-26
Sequence 26, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match 78.6%; Score 609; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.6e-56;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGPDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGPDPSTLSIYQGSANAALYDQ 60
QY 61 SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNDYDQLVTRVVT 120
DB 61 LVTRVVTHEMAHAGVNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 9
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 78.5%; Score 608; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 2.1e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGPDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGPDPQDQLVTRVVTHEMAHALQ 60
QY 61 SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNDYDQLVTRVVT 120
DB 61 SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
```

```
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 10
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match 78.2%; Score 606; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 3.4e-56;
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGPDPSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSGPDPSTLSIYQGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNDYDQLVTRVVT 120
DB 61 SPARKYDQLVTRVVTHEMAHAGQGANSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.8%; Score 603; DB 19; Length 131;
Best Local Similarity 90.1%; Pred. No. 5.9e-56;
Matches 118; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGGSGPDPSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 80
```

Db 1 GVVPQGGGNGHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 VQGGADNSTIELTQNGFRNATIDQWNAKNVDQLVTRVVTHEMAHANQTSASSVMVRQV 140
Db 61 VQGGADNSTIELTQNGFRNATIDQWNAKNSDITVGQYGGNNAALVNQTSASSVMVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12

US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-28

Query Match 77.4%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.5e-55;
Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDQWNAKNVDQLVTRVVT 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHANQTSASSVMVRQV 120
QY 121 HEMAHANQTSASSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTSASSVMVRQVFGNNATANQY 151

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.405
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16

Query Match 74.2%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.2e-53;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDQWNAKNVDQLVTRVVT 120
Db 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDQWNAKNSDITVGQYGG 120
QY 121 HEMAHANQTSASSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTSASSVMVRQVFGNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.8%; Score 533; DB 19; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.3e-48;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDQWNAKNVDQLVTRVVT 120
Db 61 TDARNSDLTITQGGGNGADVGGDDSDIDLTQRFGNSATLDQWNGKNSMTVKQFGG 120
QY 121 HEMAHANQTSASSVMVRQVFGNNATANQY 151
Db 121 GNGAIVDQTSASSVMVRQVFGNNATANQY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1998-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      68.4%; Score 530; DB 13; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.8e-48;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYYGSAANAALAQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      1 MKLLKVAATAAIVFSGSAGVAVVPQYGGGNGHGGGNSGPNSELNIIYGGGNSALAQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      61 SPARKSETTITOSGYGNGADVGCGADNSTIELTQNGFRNNATIDOWNNAKNVDQLVTRVVT 120
        :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:
DB      61 TDARNSDLTITQGGGNGADVGCGSDSDSIDLTQKFGNSATLDQWNGKNSMTVKQFEG 120
        :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:  :|||:

QY      121 HEMAHANOTASDSVMVRQVGFNNATANQY 151
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Search completed: August 2, 2004, 15:26:45
Job time : 167.9 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New:
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pap.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pap.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pap.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pap.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pap.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pap.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 530 | 68.4 | 151 | 5 | US-09-741-873C-4 |
| 2 | 452 | 58.3 | 131 | 5 | US-09-741-873C-2 |
| 3 | 91 | 11.7 | 719 | 5 | US-09-248-796A-17559 |
| 4 | 90.5 | 11.7 | 258 | 6 | US-10-425-115-300390 |
| 5 | 90.5 | 11.7 | 295 | 6 | US-10-425-115-312468 |
| 6 | 90.5 | 11.7 | 299 | 6 | US-10-170-205E-35751 |
| 7 | 90.5 | 11.7 | 520 | 6 | US-10-479-638-21 |
| 8 | 90.5 | 11.7 | 1203 | 6 | US-10-170-205E-741 |
| 9 | 90.5 | 11.7 | 1327 | 1 | PCT-US04-02338-49 |
| 10 | 89.5 | 11.5 | 256 | 6 | US-10-425-115-301334 |
| 11 | 85.5 | 11.0 | 388 | 5 | US-09-248-796A-17306 |
| 12 | 85 | 11.0 | 535 | 1 | PCT-US04-21492-88 |
| 13 | 84 | 10.8 | 376 | 6 | US-10-491-733-2 |
| 14 | 83.5 | 10.8 | 179 | 6 | US-10-425-115-346132 |
| 15 | 83.5 | 10.8 | 573 | 7 | US-60-565-632-7907 |
| 16 | 83.5 | 10.8 | 573 | 7 | US-60-579-062-7907 |
| 17 | 83 | 10.7 | 889 | 5 | US-09-952-2678-15 |
| 18 | 83 | 10.7 | 889 | 5 | US-10-872-768-15 |
| 19 | 83 | 10.7 | 889 | 6 | US-10-872-769-15 |
| 20 | 81.5 | 10.5 | 508 | 6 | US-10-425-115-285216 |
| 21 | 81.5 | 10.5 | 532 | 6 | US-10-425-115-285214 |
| 22 | 81 | 10.5 | 131 | 7 | US-60-565-632-11109 |
| 23 | 81 | 10.5 | 131 | 7 | US-60-579-062-11109 |
| 24 | 81 | 10.5 | 201 | 6 | US-10-425-115-309662 |
| 25 | 81 | 10.5 | 376 | 6 | US-10-425-115-342526 |
| 26 | 81 | 10.5 | 382 | 6 | US-10-771-241-299 |

27 81 10.5 659 1 PCT-US04-09385-4 Sequence 4, Appli
28 81 10.5 659 6 US-10-809-790-4 Sequence 4, Appli
29 80.5 10.4 246 6 US-10-854-439-511 Sequence 511 App
30 80.5 10.4 389 6 US-10-724-972A-4638 Sequence 4638, Ap
31 80.5 10.4 891 7 US-60-566-425-574 Sequence 574, App
32 80.5 10.4 891 7 US-60-576-812-606 Sequence 606, App
33 80.5 10.4 894 7 US-60-566-425-570 Sequence 570, App
34 80.5 10.4 894 7 US-60-566-425-571 Sequence 571, App
35 80.5 10.4 894 7 US-60-566-425-575 Sequence 575, App
36 80.5 10.4 894 7 US-60-566-425-578 Sequence 578, App
37 80.5 10.4 894 7 US-60-576-812-602 Sequence 602, App
38 80.5 10.4 894 7 US-60-576-812-603 Sequence 603, App
39 80.5 10.4 894 7 US-60-576-812-607 Sequence 607, App
40 80.5 10.4 894 7 US-60-576-812-610 Sequence 610, App
41 80.5 10.4 898 6 US-10-170-205E-12329 Sequence 12329, A
42 80.5 10.4 898 7 US-60-566-425-577 Sequence 577, App
43 80.5 10.4 898 7 US-60-576-812-609 Sequence 609, App
44 79 10.2 132 6 US-10-425-115-351875 Sequence 351875, A
45 79 10.2 218 5 US-09-248-796A-14423 Sequence 14423, A

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 68.4%; Score 530; DB 5; Length 151;
Best Local Similarity 69.8%; Pred.No. 1.8e-39;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQ 60
Qy 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQNAKNDLVTRVVT 120
Db 61 TDARNEDLITITQGGNGADVGQGDSSLDLTFQFGNSATLDQNGKNSMTVQFGG 120
Qy 121 HEMAHANQATASDSSVMVRQVFGNNATANY 151
Db 121 GNGAVDQATASNVVTVQVFGNNATANY 151

Matches 29; Conservative 9; Mismatches 44; Indels 5; Gaps 2;
QY 26 WGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSYGNGADVQCGA 85
DB 138 YGGGYSGGYSSG-GYAANGYGVGSGGNYSNASGGYSGS---DGYNGGAASGGYA 192
QY 86 DNSITELTQNGFRNNAIDQWNAKND 112
DB 193 NNLSSGYSNGRYNTIGSSDNGTGGYN 219

RESULT 6
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

Query Match 11.7%; Score 90.5; DB 6; Length 299;
Best Local Similarity 28.4%; Pred. No. 2.5;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;
QY 6 VAAPFAI-VVSGALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSTSS-SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVQGGADNSTIELTQNGFRNNAIDQWNAKNDQLVTRVVT 120
DB 158 GVSSTAPLPGGAAGTGTGVPVSGALRELLEACRNG-----DVSrvKRLVDA--- 204
QY 121 HEMAHANQTASD-----SSVMVRQVGF 143
DB 205 -----ANVNADWAGRKSSPLHFAAGFG 227

RESULT 7
US-10-479-638-21
; Sequence 21, Application US/10479638
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: WYO.02-0004US
; CURRENT APPLICATION NUMBER: US/10/479,638
; CURRENT FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/US02/18256
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Argiope trifasciata
US-10-479-638-21

Query Match 11.7%; Score 90.5; DB 6; Length 520;
Best Local Similarity 24.6%; Pred. No. 4.8;
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 4;

QY 15 SGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSET-TITQS 73
DB 347 AGAGAAAAASAGAGAGYGGYGVAGGS-----ISYGTSSSATSSSTASSRSGIVTSG 402
QY 74 GYGNGADVQGGADNSTIELTQNGFRNNAIDQWNAKNDQLVTRVVTHEMAHANQTASDS 133
DB 403 GYGAGAAAGAGAGAAAGAGSYG-----SISRLSSA---EAVNRVSSNIGAVASGGASAL 454
QY 134 SVMVRQVGFNNATANQY 151
DB 455 PGVISNIFSGVSSAGSY 472

RESULT 8
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

Query Match 11.7%; Score 90.5; DB 6; Length 1203;
Best Local Similarity 28.4%; Pred. No. 13;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;
QY 6 VAAPFAI-VVSGALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSTSS-SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT---QSGYNGADVQGGADNSTIELTQNGFRNNAIDQWNAKNDQLVTRVVT 120
DB 158 GVSSTAPLPGGAAGTGTGVPVSGALRELLEACRNG-----DVSrvKRLVDA--- 204
QY 121 HEMAHANQTASD-----SSVMVRQVGF 143
DB 205 -----ANVNADWAGRKSSPLHFAAGFG 227

RESULT 9
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPKAS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; TITLE OF INVENTION: USE
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: PRT

```
; ORGANISM: Homo sapiens
PCT-US04-02338-49

Query Match      11.7%; Score 90.5; DB 1; Length 1327;
Best Local Similarity 28.4%; Pred. No. 14;
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAPAAI-VVSGSALAGVVPQWGGGNNHGGGNSGPDSTLSIYQVGSNAALALQSDAR 64
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSFTSS-SSSPSPSGSLAESPEAA 157
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 65 KSETTIT---OSQYGNAGADVGGADNSTIELTQNGFRNNATIDQWNAKNYDQLVTRVVT 120
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 158 GVSSTAPLPGGAAGPTGVPVSGALRELLACRNG-----DVSRVKELVDA--- 204

QY 121 HEMAHANQIASD-----SSVMVRQVGF 143
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 10
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pep
US-10-425-115-301334

Query Match      11.5%; Score 89.5; DB 6; Length 256;
Best Local Similarity 27.9%; Pred. No. 2.6;
Matches 29; Conservative 14; Mismatches 40; Indels 21; Gaps 4;

QY 17 SALAGVVPQWGGGNNHGGGNSGPDSTLSIYQVGSNAALALQSDA---RKSETTITQ 72
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 79 SSVAG-----GGGGGGGGGTGGGGGGGGGGTSSSTRAASGPGSGNANAGKAG 133

QY 73 SGVNGAD-----VQGADNSTIEL--TQNGFRNNATID 104
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 134 GGMGGGADGAYSGAGGCVKGQGESGVALAPSSDGYNGGAAD 177

RESULT 11
US-09-248-796A-17306
; Sequence 17306, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17306
; LENGTH: 389
; TYPE: PRT
```

```
; ORGANISM: Candida albicans
US-09-248-796A-17306

Query Match      11.0%; Score 85.5; DB 5; Length 388;
Best Local Similarity 27.9%; Pred. No. 9.4;
Matches 38; Conservative 16; Mismatches 53; Indels 29; Gaps 6;

QY 30 GNHGGGNSGPDSTLSIYQVGSNAALALQSDARKSETTITQSGYGNAGADVQGGADNST 89
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 132 GNOQVGSNSYSPDT-----YGSAIGTLGVQEKTAFAVTGIHSGKIGAAA--Y 176
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 90 IELTONG-----FRNNATIDQWNAKNYDQLVTRVVT--THEMAHANQIAS-SSV 135
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 177 PELTNAGNTGLAKGTAPASTSATYGESPSADYSKSGATGVVPATYLTNSGPTGSLNTAG 236

QY 136 MVRQVGFNNATANQY 151
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 237 VVGAGFGDNSNTSSY 252

RESULT 12
PCT-US04-21492-88
; Sequence 88, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; CURRENT FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 88
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample.
; NAME/KEY: SIGNAL
; LOCATION: (1)...(27)
PCT-US04-21492-88

Query Match      11.0%; Score 85; DB 1; Length 535;
Best Local Similarity 25.0%; Pred. No. 15;
Matches 39; Conservative 19; Mismatches 62; Indels 36; Gaps 7;

QY 12 IVVSG-----SALAGVVP--QWGGGNN-----HNGG-----GNSSGPDSTLSIYQYGA 53
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 361 IVVNGTQRAENGSVNTGVWANNQCGSGNSEWLHCNGYISFGNVSGSSSSSSSSSS 420

QY 54 NAALALQSDARKSETTITQ-----SGYGNAGADVQGGADNSTIELT-QNGFRNNATIDQ 105
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 421 SSSSSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 480

QY 106 WNAKNYDQLVTRVVTTHEMAHANQIASDSSVMVRQVQ 141
   ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 481 WGVSNFTGRTISV-----TVNGSGTAVTTIG 506

RESULT 13
US-10-491-733-2
; Sequence 2, Application US/10491733
; GENERAL INFORMATION:
; APPLICANT: Syngenta Participations AG
; APPLICANT: Sainz, Manuel
; APPLICANT: Salmeron, John
```

```
; APPLICANT: Weislo, Laura J.
; TITLE OF INVENTION: Nucleic Acid Molecules from Rice Encoding Proteins for Abiotic St
; TITLE OF INVENTION: Resistance, Yield, Disease Resistance and Nutritional Quality an
; FILE REFERENCE: 60127WOPCT
; CURRENT APPLICATION NUMBER: US/10/491,733
; PRIOR APPLICATION NUMBER: 2004-04-05
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-491-733-2

Query Match      10.8%; Score 84; DB 6; Length 376;
Best Local Similarity 31.5%; Pred. No. 12;
Matches 34; Conservative 15; Mismatches 39; Indels 20; Gaps 6;

QY 17 SALAGVVPQWGGGNNHGGSSGPDSTLSIVQYGSANAALALOS--DARKSETTITQSG 74
Db 50 SALAG-----GGGG--GGGGGFGKD-----FGSMMDLLSIWTABESQAVASAG 95

QY 75 YNGADVQGGADNSTIETQNGFRNNATIDQNAKNDYDQLVTRVVTHE 122
Db 96 SAAGVGVAVGAPPTSLQ--RQSLTLPRTL---SAKTVDVVRNLVRDE 139

RESULT 14
US-10-425-115-346132
; Sequence 346132, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihue
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346132
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pep
US-10-425-115-346132

Query Match      10.8%; Score 83.5; DB 6; Length 179;
Best Local Similarity 30.4%; Pred. No. 5.8;
Matches 28; Conservative 8; Mismatches 33; Indels 23; Gaps 4;

QY 19 LAGVVPQWGGG-----NHGGGSSGPDSTLSIVQYGSANAALALQSDARKSETTI 70
Db 64 LDGLLGLGGLGGGLDGLLGLTGGGGKKNQADSGNAQDSGNAQ-----QEDSGNAQ--- 116

QY 71 TQSGYNGADVQGGADNSTIETQNGFRNNAT 102
Db 117 EESGNNQAGAGAGA-----ENGAAANGT 140

RESULT 15
US-60-565-632-7907
; Sequence 7907, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
```

```
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; TITLE OF INVENTION: Compositions Thereof
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 7907
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
US-60-565-632-7907

Query Match      10.8%; Score 83.5; DB 7; Length 573;
Best Local Similarity 27.8%; Pred. No. 22;
Matches 45; Conservative 15; Mismatches 73; Indels 29; Gaps 7;

QY 10 AAIWVSGSALAGVVPQWGGGNNHGGSSGPDSTLSIVQYGSANAALA-----LQSD 62
Db 155 AADNNGSADAA-----QGNDNRAAANNANADAQTDAQAQ-GSANEANAENNANADAQND 208

QY 63 ARKSETTITQSGYNGADVQGGADN-STIELTON-----GFRNNATID-----QWNK 109
Db 209 AAQANDNGAAEAENNGNADAAQGGTDNEAAEENSGNENGNGTGAENNNANADAQTQDVAQGSTNEA 268

QY 110 NYDQLVTRVVTHEMAHANQ---TASDSSVMVRQVGFNNATA 148
Db 269 NAENNANADVQNDAAQAQANENGAAEENSGNADAAQGGTDNGAAA 310

Search completed: August 2, 2004, 15:29:54
Job time : 17.8 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)

1545.204 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 690 | 89.0 | 151 | 2 JC6039 | fimbrin protein ag |
| 2 | 690 | 89.0 | 151 | 2 A10635 | major curlin chain |
| 3 | 533 | 68.8 | 151 | 2 S70788 | curlin protein csg |
| 4 | 511.5 | 66.0 | 152 | 2 D90806 | curlin major subun |
| 5 | 511.5 | 66.0 | 152 | 2 H85665 | hypothetical prote |
| 6 | 101 | 13.0 | 409 | 2 T20847 | hypothetical prote |
| 7 | 99 | 12.8 | 1748 | 2 S42136 | cn3B protein - tet |
| 8 | 98.5 | 12.7 | 151 | 2 JC6040 | fimbrin protein ag |
| 9 | 98.5 | 12.7 | 151 | 2 AH0635 | nucleation compone |
| 10 | 96.5 | 12.5 | 151 | 2 S70787 | curlin nuclearor p |
| 11 | 96.5 | 12.5 | 151 | 2 C90806 | minor curlin subun |
| 12 | 96.5 | 12.5 | 151 | 2 G85665 | curlin minor chain |
| 13 | 93.5 | 12.1 | 552 | 2 T070604 | probable PPE prote |
| 14 | 92.5 | 11.9 | 440 | 2 AD1539 | probable sugar ABC |
| 15 | 92.5 | 11.9 | 1028 | 2 A56038 | DNA-binding protei |
| 16 | 92.5 | 11.9 | 1213 | 2 S16356 | ovo protein - frui |
| 17 | 90 | 11.6 | 145 | 2 AD3143 | conserved hypothet |
| 18 | 90 | 11.6 | 145 | 2 H98144 | hypothetical prote |
| 19 | 90 | 11.6 | 347 | 2 H39112 | merozoite 45K surr |
| 20 | 89.5 | 11.5 | 256 | 2 T03371 | glycine-rich prote |
| 21 | 89 | 11.5 | 262 | 2 S00275 | tail fiber protein |
| 22 | 89 | 11.5 | 573 | 2 C86266 | F3F19.21 protein - |
| 23 | 89 | 11.5 | 2174 | 2 E95965 | hypothetical glyci |
| 24 | 88.5 | 11.4 | 221 | 2 A47369 | RNA-binding protei |
| 25 | 87 | 11.2 | 590 | 1 A45621 | leishmanolysin (EC |
| 26 | 87 | 11.2 | 599 | 2 B42049 | leishmanolysin (EC |
| 27 | 87 | 11.2 | 599 | 2 A44951 | leishmanolysin (EC |
| 28 | 87 | 11.2 | 602 | 1 PL0221 | leishmanolysin (EC |
| 29 | 87 | 11.2 | 646 | 1 S19916 | leishmanolysin (EC |

ALIGNMENTS

RESULT 1

JC6039

fimbrin protein agfA precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C:Date: 31-Dec-1996 #sequence,revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: JC6039; PC6015; A44898

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.

A:Reference number: JC6039; MUID:96146512; PMID:8550497

A:Accession: JC6039

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714

A:Accession: PC6015

A:Molecule type: protein

A:Residues: 21-52 <CO2>

A:Experimental source: strain 27655-3b

A:Note: the authors translated the codon ACG for residue 44 as Ile

R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A:Title: Purification and characterization of thin, aggregative fimbrinae from Salmonella

A:Reference number: A44898; MUID:9310586; PMID:1677357

A:Contents: 27655

A:Accession: A44898

A>Status: preliminary

A:Molecule type: protein

A:Residues: 21-33 <CO3>

A:Note: sequence extracted from NCBI backbone (NCBIP:45936)

C:Genetics:

A:Gene: agfA

C:Function:

A:Description: major component of thin aggregative fimbrinae

A:Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbrina

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match 89.0%; Score 690; DB 2; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.7e-51;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ | 60 |
| DB | 1 | MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ | 60 |
| QY | 61 | SDARKSETTTTQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKYDQAVRWVT | 120 |
| DB | 61 | SDARKSETTTTQSGYNGADVGQGDADNSTIELTQNGFRNNATIDQWNAKYDQAVRWVT | 120 |
| QY | 121 | HEMAFANQTAQSDSSVMVRQVGFNNATANQY | 151 |

CT147 hypothetical
conserved hypothet
probable disease r
probable PPE prote
probable secreted
protein C05B5.3 [i
protein kinase sgg
hypothetical prote
hypothetical prote
probable sugar ABC
hypothetical prote
hemolysin (importe
leishmanolysin (EC
hypothetical prote
probable outer mem

30 87 11.2 1537 2 F86509
31 87 11.2 1537 2 C81558
32 86 11.1 447 2 G84687
33 86 11.1 582 2 F70675
34 85.5 11.0 438 2 T35789
35 85 11.0 401 2 C88571
36 85 11.0 575 2 S35327
37 85 11.0 967 2 S66852
38 84.5 10.9 340 2 A83401
39 84 10.8 439 2 AC1182
40 84 10.8 764 2 H71607
41 83.5 10.8 423 2 T19581
42 83.5 10.8 1635 2 AI0452
43 83 10.7 639 2 C42049
44 82.5 10.6 343 2 T05221
45 82 10.6 382 2 F90892

Db 121 NNAALVNTASDSSVMVRQVFGNNATANOY 151

RESULT 2

AI0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhimurium
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY181

Query Match 89.0%; Score 690; DB 2; Length 151;
Best Local Similarity 91.4%; Pred. No. 4.7e-51;
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120

DB 61 SDARKSETTITQSGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120

QY 121 HEMAHANQASDSSVMVRQVFGNNATANOY 151

DB 121 NNAALVNTASDSSVMVRQVFGNNATANOY 151

RESULT 3

S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34559
R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and C

A:Reference number: S70788

A:Accession: S70788

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAAG2282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A6720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;
A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V',8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OUS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'R'ORDSGWLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 68.8%; Score 533; DB 2; Length 151;
Best Local Similarity 70.2%; Pred. No. 7.1e-38;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 120

DB 61 TDARNSDLTITQHGNGGADVGQSDSSIDLTFQGFNSATLDQWNGKNSMTVKQFGG 120

QY 121 HEMAHANQASDSSVMVRQVFGNNATANOY 151

DB 121 GNGAIVDQATASNSVNVTVQVFGNNATANOY 151

RESULT 4

D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 050952

C:Genetics:

A:Gene: ECs1420

Query Match 66.0%; Score 511.5; DB 2; Length 152;

Best Local Similarity 68.4%; Pred. No. 4.6e-36;

Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAL 59

DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAL 60

QY 60 QSDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNATIDQWNAKNDQLVTRVVT 119

Db 61 QADAENSLTITQHGGNGADVGGSDSSIDLTRQFGNSATLDFQWNGKDSHMTVKQFG 120
QY 120 THEMAHANOTASDSSVMVQVGFQGNATANQY 151
Db 121 GGNGAAMDQATASNTVNTVQVGFQGNATAHQY 152

RESULT 5

H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Fejnara, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:CROSS-references: GB:AE005174; NID:G12514574; PIDN:AA655788.1; GSPDB:GN00145; UWGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match 66.0%; Score 511.5; DB 2; Length 152;
Best Local Similarity 68.4%; Pred. No. 4.6e-36;
Matches 104; Conservative 19; Mismatches 26; Indels 1; Gaps 1;
QY 1 MKLLKVAFAATVVGSSALAGVVPQW-GGGGNHGGNSGPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAIAAIVFGSALAGVVPQYGGGNGHGGNSGPNSELNIIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVQGGADNSTIELTQGFNRNATIDQWNAKNDQLVTRV 119
Db 61 QADAENSLTITQHGGNGADVGGSDSSIDLTRQFGNSATLDFQWNGKDSHMTVKQFG 120
QY 120 THEMAHANOTASDSSVMVQVGFQGNATANQY 151
Db 121 GGNGAAMDQATASNTVNTVQVGFQGNATAHQY 152

RESULT 6

T20847
hypothetical protein Fl3E9.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Aug-2002
C:Accession: T20847
R:McMurray, A.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19332
A:Accession: T20847
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-409 <WIL>
A:CROSS-references: EMBL:Z69383; PIDN:CAA93412.1; GSPDB:GN00022; CESP:Fl3E9.4
A:Experimental source: clone Fl3E9
C:Genetics:
A:Gene: CESP:Fl3E9.4
A:Map position: 4
A:Introns: 32/1; 275/3; 337/3
C:Superfamily: loricrin

Query Match 13.0%; Score 101; DB 2; Length 409;
Best Local Similarity 23.3%; Pred. No. 0.4;
Matches 37; Conservative 26; Mismatches 66; Indels 30; Gaps 5;
QY 15 SGLSALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG 74
Db 134 SQQASGNSVNSFGQGGYQNGFGGQSGFSGQSGWGSNSLSNSGNNNGQ--SSSG 191

QY 75 YGN--GADVQGGADNSTIE---LTQNGFRNN-----ATIDOWNAK 109
Db 192 YQNNQGRHQCGGCGSHSSSSSVMSNNGYSNGGNNNGTPTSFLLNNVSSAAQDYINIV 251
QY 110 NYDQLVTRVTHEMAHANOTASDSSVMVQVGFQGNATA 148
Db 252 NNSKLTNTQINEQ---ASNWASANSVQAQYIQTETNRS 287

RESULT 7

S42136
cnjB protein - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C>Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
submitted to the EMBL Data Library, October 1992
A:Reference number: S42136
A:Accession: S42136
A:Molecule type: DNA
A:Residues: 1-1748 <TAY>
A:CROSS-references: EMBL:L03710; NID:G161751; PID:G161752
R:Taylor, F.M.; Martindale, D.W.
Nucleic Acids Res. 21, 4610-4614, 1993
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by
A:Reference number: S42135; MUID:94051569; PMID:8233798
A:Accession: S42135
A:Molecule type: DNA
A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-
A:CROSS-references: EMBL:L03710
R:Martindale, D.W.; Taylor, F.M.
Nucleic Acids Res. 16, 2189-2201, 1988
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.
A:Reference number: S03650; MUID:88189811; PMID:3357771
A:Accession: S03650
A:Molecule type: DNA
A:Residues: 236-250; '1', 252-255, 'N', 257-773 <MAR>
A:CROSS-references: EMBL:X06462
C:Genetics:
A:Gene: cnjB

A:Genetic code: SGC5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3;
C:Keywords: zinc finger
F:1164-1450/Region: Glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query Match 12.8%; Score 99; DB 2; Length 1748;
Best Local Similarity 28.0%; Pred. No. 3.1;
Matches 37; Conservative 16; Mismatches 31; Indels 48; Gaps 6;

QY 25 QWGGGNGHGG---GNSSGPDSTLSIYQYGSANAALALQSDARKSETTIT---QSGYGN 77
Db 1640 QFGGGGNSGQSGWGTSSGSDWN-----CQSNVQESTTSSGGWSSGSGN 1685
QY 78 GADVQGGADNSTIELTQNGFRNATIDQWNAKNDQLVTRVTHEMAHANOTASDSSVMV 137
Db 1686 QTGGGWSNDN-----QQQNTGGGWSGSSN-----SNQTNSS--- 1722
QY 138 QVGFQGNATAN 149
Db 1723 ----WGSNNQAS 1730

RESULT 8
JC6040

fimbrin protein agfB precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: J06040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:G1184712; PIDN:AA43598.1; PID:G1184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
C:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbrin protein agfB #status predicted <MAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.21;
Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94

Db 14 GAPGATATNYDLARSEYFVNFVNEISKSFNQAAIGQVGTDSARVQEGSKLLSVISQ 73

QY 95 NGRFNATIDQNAKND-QLTRVVVTHEMAHANTASDSSVMVQVGFQGNNTANQY 151

Db 74 EGGNNRAKVDQ--AGNYPAYIEQTGNANDASISQAYGNSAAIIQKSGNKNANTQY 129

RESULT 9

AH0635

nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AH0635

R:Farhail, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AH0635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:G16502314; GSPDB:GN00176

C:Genetics:

A:Gene: STY1180

Query Match 12.7%; Score 98.5; DB 2; Length 151;

Best Local Similarity 28.8%; Pred. No. 0.21;

Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94

Db 14 GAPGATATNYDLARSEYFVNFVNEISKSFNQAAIGQVGTDSARVQEGSKLLSVISQ 73

QY 95 NGRFNATIDQNAKND-QLTRVVVTHEMAHANTASDSSVMVQVGFQGNNTANQY 151

Db 74 EGGNNRAKVDQ--AGNYPAYIEQTGNANDASISQAYGNSAAIIQKSGNKNANTQY 129

RESULT 10

S70787

curlin nucleator protein csb precursor - Escherichia coli (strain K-12)

N:Alternate names: csbB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csb operons is required for production of fibronectin- and C

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70787

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:G1147558; PIDN:CAA62281.1; PID:G1147563

A:Experimental source: strain K12, substrain W310

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:G1787265; PIDN:AA474125.1; PID:G1787278

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csbB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 12.5%; Score 96.5; DB 2; Length 151;

Best Local Similarity 25.6%; Pred. No. 0.31;

Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQGDADNSTIELTQGNFNATIDQ 105

Db 14 GAPGATAAGYDLARSEYFVNFVNEISKSFNQAAIGQAGTNSAGLRGGSKLLAVVQAQ 73

QY 106 WNAKYNQDLVTRVVTHEMAHANO--TASDSSV-----MVRQVGFQGNNTANQY 151

Db 74 EGSSNRKAK-IDQTGDYNLAYIDQAGSANDASISQAYGNTAMTIQKSGNKNANTQY 129

RESULT 11

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subse

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: C90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <HAY>

A:Cross-references: GB:BA000007; PIDN:BAB34842.1; PID:G13360879; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs1419

Query Match 12.5%; Score 96.5; DB 2; Length 151;

Best Local Similarity 25.6%; Pred. No. 0.31;

Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

| | | | | | | | | | | | | | | |
|----|-----|-------|------|------|------|-------|--------|-------|------|--------|-------|--------|-------|-------|
| Db | 59 | LQNA | AAAA | YIMS | AGSG | ----- | GGGTG | NGGGG | AGSG | PGG | PSANS | GGGGGG | ----- | 104 |
| Qy | 63 | ARKS | ETIT | QSGY | NGAD | VGG | ADNST | IELT | QNGF | RNNATI | QDWN | AKNYD | QIVTR | VVTHE |
| Db | 105 | ----- | GGNG | YINC | GGVG | -GP | NNS--- | LDG | NNLL | NFAS | VS | NYNES | ----- | SKFH |
| Qy | 123 | MAHA | ---- | NOT | ASD | SSV | VRQ | VFG | NN | NATA | 148 | | | |
| Db | 148 | HHQH | NNNN | NNNG | GGQT | SM | MG | HP | FG | GN | PSA | 177 | | |

Search completed: August 2, 2004, 14:56:24
 Job time : 9.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds

(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSVNVVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 690 | 89.0 | 151 | 1 CSGA_SALTY | P55225 salmonella |
| 2 | 533 | 68.8 | 151 | 1 CSGA_ECOLI | P28307 escherichia |
| 3 | 511.5 | 66.0 | 152 | 1 CSGA_ECO57 | Q93u24 escherichia |
| 4 | 98.5 | 12.7 | 151 | 1 CSGB_SALTY | Q93u24 salmonella |
| 5 | 98.5 | 12.7 | 151 | 1 CSGB_SALTY | P55226 salmonella |
| 6 | 96.5 | 12.5 | 151 | 1 CSGB_ECOLI | P39828 escherichia |
| 7 | 92.5 | 11.9 | 1028 | 1 OVO_DROME | P51521 drosophila |
| 8 | 90.5 | 11.7 | 1327 | 1 TNK1_HUMAN | Q95271 homo sapien |
| 9 | 90 | 11.6 | 347 | 1 MSA2_PLAF2 | Q03646 plasmodium |
| 10 | 89 | 11.5 | 262 | 1 VG38_BPT2 | P07875 bacterioph |
| 11 | 87 | 11.2 | 590 | 1 GP63_LEIDO | P23223 leishmania |
| 12 | 87 | 11.2 | 599 | 1 GP63_LEICH | P15706 leishmania |
| 13 | 87 | 11.2 | 602 | 1 GP63_LEIMA | P08148 leishmania |
| 14 | 87 | 11.2 | 646 | 1 GP63_LEIME | P43150 leishmania |
| 15 | 87 | 11.2 | 1656 | 1 OMPB_RICUA | P06653 x outer mem |
| 16 | 85 | 11.0 | 401 | 1 YK03_CABEL | P34291 caenorhabdi |
| 17 | 82 | 10.6 | 678 | 1 YF48_MYCTU | Q10778 mycobacteri |
| 18 | 82 | 10.6 | 1093 | 1 PER_BROWI | Q03297 drosophila |
| 19 | 82 | 10.6 | 1115 | 1 TB22_CHLRE | Q8xyp3 chlamydonon |
| 20 | 81.5 | 10.5 | 306 | 1 HMXD_DROPS | P20822 drosophila |
| 21 | 81.5 | 10.5 | 311 | 1 HMXD_PSESP | P07662 pseudomonas |
| 22 | 81.5 | 10.5 | 392 | 1 HME1_HUMAN | Q25925 homo sapien |
| 23 | 81 | 10.5 | 165 | 1 GRP1_ORYSA | Q05074 oryza sativ |
| 24 | 80.5 | 10.4 | 548 | 1 CEAK_ECOLI | Q47502 escherichia |
| 25 | 80.5 | 10.4 | 720 | 1 G7AK_BREDI | Q915d6 brevundinon |
| 26 | 80.5 | 10.4 | 894 | 1 ILF3_HUMAN | Q12906 h interleuk |
| 27 | 80 | 10.3 | 427 | 1 CFIA_DROME | P18241 drosophila |
| 28 | 80 | 10.3 | 491 | 1 YK98_MYCTU | Q10707 mycobacteri |
| 29 | 79.5 | 10.3 | 172 | 1 CH18_DROME | P07184 drosophila |
| 30 | 79.5 | 10.3 | 760 | 1 YBIL_ECOLI | P75780 escherichia |
| 31 | 79.5 | 10.3 | 1034 | 1 ICEN_PANAN | Q47879 pantoea ana |
| 32 | 79.5 | 10.3 | 1258 | 1 ICEN_ERWHE | P16239 erwinia her |
| 33 | 79.5 | 10.3 | 1322 | 1 ICFA_PANAN | P20469 pantoea ana |

RESULT 1

CSGA_SALTY

ID CSGA_SALTY STANDARD; PRT; 151 AA.

AC P55225;

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Major curlin subunit precursor (Fimbrin SFE17).

GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.

OS Salmonella typhimurium,

OS Salmonella typhi, and

OS Salmonella enteritidis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_TaxID=602, 601, 592;

RN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=SR-11;

RX MEDLINE=98117058; PubMed=9457880;

RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;

RT "Curli fibers are highly conserved between Salmonella typhimurium and

Escherichia coli with respect to operon structure and regulation.";

RL J. Bacteriol. 180:722-731(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,

Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";

RL Nature 413:852-856(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhi; STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,

Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

Feltwell T., Hamlin N., Haque R.M., Hien T.T., Holroyd S., Jageis K.,

Krogh A., Larsen T.S., Leather S., Moule S., O'Gea P., Parry C.,

Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,

Whitehead S., Barrall B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella

enterica serovar Typhi CT18.";

RL Nature 413:848-852(2001).

RN [4]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22531367; PubMed=12644504;

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,

Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

"Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18";
 J. Bacteriol. 185:2330-2337(2003).
 [5]
 RN SEQUENCE FROM N.A.
 RP SPECIES-S. enteritidis; STRAIN=27655-3B;
 RC MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae";
 J. Clin. Microbiol. 178:662-667(1996).
 [6]
 RN SEQUENCE OF 21-151 FROM N.A.
 RP SPECIES-S. enteritidis; STRAIN=27655-3B;
 RC MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 Munro C.K., Kay W.W., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella enteritidis* targeting agfA,
 the structural gene for thin, aggregative fimbriae";
 J. Clin. Microbiol. 31:2263-2273(1993).
 [7]
 RN SEQUENCE OF 21-33.
 RP SPECIES-S. enteritidis; STRAIN=27655-3B;
 RC MEDLINE=9130586; PubMed=1677357;
 RA Collinson S.K., Emedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis";
 J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AJ002301; CAA05317.1; -;
 DR EMBL; AF008749; AAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; AS016840; AAC69399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR FIC; J06039; J06039.
 DR StyGene; SG10608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMVQVGFNNATANYQ -> DSVTQVAS (IN
 REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 89.0%; Score 690; DB 1; Length 151;
 Best Local Similarity 91.4%; Pred. No. 1.4e-51;
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGNGSGPDSITLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGNGSGPDSITLSIYQYGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDOWNAKNYDQVTVRVVT 120
 DB 61 SPARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
 QY 121 HEMAHANQASDSSVMVQVGFNNATANYQ 151
 DB 121 NNAALVNQASDSSVMVQVGFNNATANYQ 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID _CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli";
 J. Mol. Microbiol. 7:523-536(1993).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / MC4100;
 RX MEDLINE=96414469; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 K-12";
 J. Mol. Microbiol. 18:661-670(1995).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=92798503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12";
 Science 277:1233-1238(1997).
 [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map";
 DNA Res. 3:137-155(1996).
 [5]
 RN SEQUENCE OF 21-40.
 RP STRAIN=K12 / YMEL;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The curli protein activates cryptic genes for curli formation and
 fibronectin binding in *Escherichia coli* HB101";
 J. Mol. Microbiol. 6:2443-2452(1992).
 [6]
 RN SEQUENCE OF 21-31.
 RP MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis";
 J. Bacteriol. 173:4773-4781(1991).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.

```

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAR23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR EMBL; S70788; S70788.
DR ECoGene; EG11489; csga.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7
FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
SQ
Query Match 68.8%; Score 533; DB 1; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.4e-38;
Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETITQSGYNGADYVGGADNSTIELTQNGFNATIDQWNAKYDQVLTVRV 120
Dd 61 TDARNSDLTITQHGNGGADYVGGSDSDITLQKFGNSATLDQWNGKNSMTVKQFG 120
QY 121 HEMAHANTASDSSVMVRQVGFNNATANY 151
Dd 121 GNGAAVDOTASNSVTVVTVQVGFNNATAHQY 151
RESULT 3
CSGA_ECO57 ID CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECG1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csd promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533(2001).

```

```

[3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
RX MEDLINE=21156231; PubMed=11256796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT SEQUENCE 152 AA; 15099 MW; EE2D294DDE91243 CRC64;
SQ
Query Match 66.0%; Score 511.5; DB 1; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.5e-36;
Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYQYGSANAALAL 59
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGHNGSGSPDSTLSIYQYGSANAALAL 60
QY 60 QSDARKSETITQSGYNGADYVGGADNSTIELTQNGFNATIDQWNAKYDQVLTVRV 119
Dd 61 QADARNSDLTITQHGNGGADYVGGSDSDITLQKFGNSATLDQWNGKNSMTVKQFG 120
QY 120 THEMAHANTASDSSVMVRQVGFNNATANY 151
Dd 121 GNGAAVDOTASNSVTVVTVQVGFNNATAHQY 152
RESULT 4
CSGB_SALTI ID CSGB_SALTI STANDARD; PRT; 151 AA.
AC Q9Z7M3;
DC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

```

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.,
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12645504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627269; CAD08267.1; -;
CC EMBL; AE016840; AAC069400.1; -;
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SEQUENCE 151 AA; 16254 MW; 161C543268573495 CRC64;
Query Match 12.7%; Score 98.5; DB 1; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.099;
Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;
QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94
Db 14 GAGCIATATNYDLARSEYFANVELSKSFNQAAIIQVGTDSARVQEGSKLSVISQ 73
QY 95 NGRFNATIDQWAKNYD-QLVTRVVTHEMAHANOTASDSSVMVROVGFNNATANQY 151
Db 74 EGENNRKVDQ--AGNYNFAYIEQTGNANDASISQSGVNSAAIIQKSGNKNATQY 129
RESULT 5
CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli" fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.,
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12645504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627269; CAD08267.1; -;
CC EMBL; AE016840; AAC069400.1; -;
CC Fimbrin; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SEQUENCE 151 AA; 16254 MW; 161C543268573495 CRC64;
Query Match 12.7%; Score 98.5; DB 1; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.099;
Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;
QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVQ-GADNST-----IELTQ 94
Db 14 GAGCIATATNYDLARSEYFANVELSKSFNQAAIIQVGTDSARVQEGSKLSVISQ 73
QY 95 NGRFNATIDQWAKNYD-QLVTRVVTHEMAHANOTASDSSVMVROVGFNNATANQY 151
Db 74 EGENNRKVDQ--AGNYNFAYIEQTGNANDASISQSGVNSAAIIQKSGNKNATQY 129
RESULT 5
CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli" fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).

RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 RT fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 RT K-12";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-Kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map";
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimmalanta E.T., Potamouzis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258786;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuwara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S.;
 RT "Sigma S-dependent growth-phase induction of the *csgBA* promoter in
 RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
 RT of the nucleoid-associated protein H-NS";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X90754; CAA62281.1; --
 DR EMBL; AE000205; AAC74125.1; --
 DR EMBL; D90741; BAA35831.1; --
 DR EMBL; AE005315; AAG55787.1; --
 DR EMBL; AP002554; BAB34842.1; --
 DR PIR; C90806; C90806
 DR PIR; G85665; G85665
 DR PIR; S70787; S70787
 DR EcoGene; EGI2621; csgB.
 KW Fibria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 15682 MW; B18D266B964014B8 CRC64;
 Query Match 12.5%; Score 96.5; DB 1; Length 151;
 Best Local Similarity 25.6%; Pred. No. 0.15;
 Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;
 QY 51 GSANAALQSDARKS-----TTTQSGYGVGADVGQADNSTIETLTQNGFRNNATIDQ 105
 Db 14 GAPGIAAAGYDLANSEYFAVNELSKSFSNQAAIICQAGTNNSAQRQSGSKLLAVVAQ 73
 QY 106 WNAKNYDQVTRVVTHEMAHANQ--TASDSV-----MVRQVGFNNATANQY 151
 Db 74 EGSSNRRAK-IDQTGYNLAYIDQAGSANDASISQAGYNTAMTIQSGNGKANITQY 129
 RESULT 7
 OVO DROME
 ID - OVO DROME STANDARD; PRT; 1028 AA.
 AC P51521; Q9XZU4;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE OVO protein (Shaven baby protein).
 GN OVO OR SVB.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=95021209; PubMed=7935398;
 RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
 RT "Multiple products from the shavenbaby-ovo gene region of *Drosophila*
 RT melanogaster: relationship to genetic complexity";
 RL Mol. Cell. Biol. 14:6809-6818(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=91293102; PubMed=1712294;
 RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
 RT "The ovo gene of *Drosophila* encodes a zinc finger protein required
 RT for female germ line development";
 RL EMBO J. 10:2259-2266(1991).
 CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
 CC LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (potential).
 CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMIUM AND
 CC ACCUMULATES IN NURSE CELLS DURING OÖGENESIS. STORED IN THE EGG,
 CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
 CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
 CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U11383; AAB60216.1; ALT_SEQ.
DR EMBL; X59772; CAB36921.1; ALT_SEQ.
DR PIR; A56038; A56038.
DR HSSP; P07248; 2ADR.
DR TRANSFAC; T00669; --
DR FlyBase; FBgn0003028; ovo.
DR InterPro; IPR007087; znf_C2H2.
DR Pfam; PF000096; zf-C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 62 66 POLY-ALA.
FT DOMAIN 72 77 POLY-GLY.
FT DOMAIN 80 85 POLY-GLY.
FT DOMAIN 98 108 POLY-GLY.
FT DOMAIN 144 152 POLY-HIS.
FT DOMAIN 153 159 POLY-ASN.
FT DOMAIN 336 339 POLY-GLN.
FT DOMAIN 347 353 POLY-GLN.
FT DOMAIN 357 361 POLY-GLN.
FT DOMAIN 410 414 POLY-GLN.
FT DOMAIN 418 422 POLY-GLN.
FT DOMAIN 426 432 POLY-GLN.
FT DOMAIN 445 453 POLY-GLN.
FT DOMAIN 456 459 POLY-GLN.
FT DOMAIN 466 474 POLY-GLN.
FT DOMAIN 497 517 POLY-ALA.
FT DOMAIN 524 529 POLY-SER.
FT DOMAIN 549 558 POLY-ALA.
FT DOMAIN 639 651 POLY-ALA.
FT DOMAIN 717 725 POLY-ALA.
FT DOMAIN 797 802 POLY-GLN.
FT DOMAIN 820 832 POLY-GLN.
FT DOMAIN 826 832 POLY-GLN.
FT ZN_FING 874 896 C2H2-TYPE 1.
FT ZN_FING 902 924 C2H2-TYPE 2.
FT ZN_FING 930 953 C2H2-TYPE 3.
FT ZN_FING 969 992 C2H2-TYPE 4.
FT ZN_FING 647 647 A -> R (IN REF. 2).
SQ SEQUENCE 1028 AA; D70688B2BC0F6F77 CRC64;

Query Match 11.9%; Score 92.5; DB 1; Length 1028;
Best Local Similarity 26.7%; Pred. No. 2.8;
Matches 40; Conservative 14; Mismatches 61; Indels 35; Gaps 6;
QY 3 LLKVAFAAIVSGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSD 62
Db 59 LQNAAAAYIMSAGS-----GGCTGCGGAGSGGPGPSANSNGGGGG----- 104
QY 63 ARKSETITSGYGNAGVCGADNSTIBLTQGFNNATIDQWAKNYDQLVTRVTHE 122
Db 105 -----GGGYINGCGVG-GPNNS---LDGNLLNFASVSNYESN-----SKPHNH 147
QY 123 MAHA---NQTASDSSVMVRQVGFNNATA 148
Db 148 HHQHNNNNNGGQISMCHPFGGNPSA 177

RESULT 8
ID_TNKL_HUMAN STANDARD; PRT; 1327 AA.
AC O95271; O95272;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-
DE interacting ankyrin-related ADP-ribose polymerase).
CN CN TNKS OR TNKS1 OR TIN1 OR TIN1 OR PARP1.
OS OS Homo sapiens (Human).
OC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutharia; Primates; Catarrhini; Hominoidea; Homo.
OX OX NCBI_TaxID=9606;
RN RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP RP TISSUE=Testis;
RX RX MEDLINE=99040105; PubMed=9822378;
RA RA Smith S., Giriat L., Schmitt A., de Lange T.;
RT RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
RL RL Science 282:1484-1487(1998).
RN RN SUBCELLULAR LOCATION.
RP RP MEDLINE=99454782; PubMed=10523501;
RX RX Smith S., de Lange T.;
RA RA "Cell cycle dependent localization of the telomeric PARP, tankyrase,
RT RT to nuclear pore complexes and centrosomes.";
RL RL J. Cell Sci. 112:3649-3656(1999).
RN RN FUNCTION AND PHOSPHORYLATION.
RP RP MEDLINE=20556282; PubMed=10988299;
RX RX Chi N.-W., Lodish H.F.;
RT RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
RL RL substrate that interacts with IRAP in GLUT4 vesicles.";
RN RN J. Biol. Chem. 275:38437-38444(2000).
RP RP FUNCTION AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
RX RX MEDLINE=21602874; PubMed=11739745;
RA RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
RT RT "Role for the related poly(ADP-ribose) polymerases tankyrase 1 and 2
RL RL at human telomeres.";
RN RN Mol. Cell. Biol. 22:332-342(2002).
RP RP FUNCTION: May regulate vesicle trafficking and modulate the
RX RX subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
RA RA activity and can modify TRF1, and thereby contribute to the
RT RT regulation of telomere length.
RL RL CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose}(N)-acceptor =
RN RN nicotinamide + {ADP-D-ribose}(N+1)-acceptor.
RP RP SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
RX RX the cytoplasmic domain of LNPEP/Oase in SLC2A4/GLUT4-vesicles.
RA RA Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
RT RT SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
RL RL with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
RN RN also found at nuclear pore complexes and around the pericentriolar
RP RP matrix of mitotic centrosomes. During interphase, a small fraction
RX RX of TNKS is found in the nucleus, associated with TRF1.
RA RA ALTERNATIVE PRODUCTS:
RT RT Event=Alternative splicing; Named isoforms=2;
RL RL Name=1;
RN RN IsoId=O95271-1; Sequence=Displayed;
RP RP Name=2;
RX RX IsoId=O95271-2; Sequence=VSP_004538, VSP_004539;
RA RA Note=No experimental confirmation available;
RT RT TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
RL RL PTM: Upon insulin-stimulation, phosphorylated on serine residues
RN RN by MAPK kinases.
RP RP PTM: ADP-ribosylated (-auto).
RX RX SIMILARITY: Belongs to the PARP family.
RA RA SIMILARITY: Contains 15 ANK repeats.
RT RT SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
RL RL This SWISS-PROT entry is copyright. It is produced through a collaboration
RN RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP RP the European Bioinformatics Institute. There are no restrictions on its
RX RX use by non-profit institutions as long as its content is in no way
RA RA modified and this statement is not removed. Usage by and for commercial
RT RT entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
RL RL or send an email to license@isb-sib.ch).
RN RN -----
RP RP EMBL; AF082556; AAC79841.1; --


```
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
DE 38.
GN Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10664;
RN [1]_SEQUENCE FROM N.A.
RP MEDLINE=87283911; PubMed=3302276;
RX Riede I., Drexler K., Eschbach M.L., Henning U.;
RA "DNA sequence of genes 38 encoding a receptor-recognizing protein of
RT bacteriophages T2, K3 and of K3 host range mutants.";
RL J. Mol. Biol. 194:31-39(1987).
CC -1- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as
CC the phage recognition site for the cellular receptor.
CC -1- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
CC AS RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05312; CAA28935.1; -.
DR PIR; S00275; S00275.
DR InterPro; IPR007932; Tail_fibre_GP38.
DR Pfam; PF05268; GP38; 1.
KW Fiber protein; Phage recognition.
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;

Query Match 11.5%; Score 89; DB 1; Length 262;
Best Local Similarity 34.4%; Pred. No. 1.2;
Matches 32; Conservative 9; Mismatches 38; Indels 14; Gaps 5;

QY 27 GGGGNGGNGSGSPDSTLSIVQYGSANAALQSDARKSETTITQSYGNGADVGGQAD 86
Db 175 GGGGNGGNGSGSPDSTLSIVQYGSANAALQSDARKSETTITQSYGNGADVGGQAD 86
QY 87 ---NSLTETGTFGRNATIDQWNAKVDQLVT 116
Db 226 RGWKNVYVSEGGAGAAVTG--NAPNQNVT 256

RESULT 11
GP63 LEIDO STANDARD; PRT; 590 AA.
ID GP63 LEIDO STANDARD; PRT; 599 AA.
AC P23223; 1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=LV9;
RC MEDLINE=92107220; PubMed=1762629;
RX Webb J.R., Button L.L., McMaster R.W.;
RA "Heterogeneity of the genes encoding the major surface glycoprotein
RT of Leishmania donovani.";
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -1- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
```

RA Miller R.A., Reed S.G., Parsons M.;
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
 RL Arg-Gly-Asp sequence.";
 RL Mol. Biochem. Parasitol. 39:267-274(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112918; PubMed=1370484;
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
 RA Wilson M.E.;
 RT "Three distinct RNAs for the surface protease gp63 are differentially
 RT expressed during development of Leishmania donovani chagasi
 RT promastigotes to an infectious form.";
 RL J. Biol. Chem. 267:1888-1895(1992).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M80672; AAA29238.1; -.
 DR EMBL; M28527; AAA29235.1; -.
 DR PIR; A44951; A44951.
 DR HSSP; P08148; 1LML.
 DR MEROPS; M08.001; -.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001577; Peptidase_M8.
 DR Pfam; PF01457; Peptidase_M8; 1.
 DR PRINTS; PR00782; LSHMANOLYSIN.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
 FT SIGNAL 1 39
 FT PROPEP 40 97
 FT CHAIN 98 574
 FT PROPEP 575 599
 FT METAL 261 261
 FT ACT_SITE 262 262
 FT METAL 265 265
 FT METAL 331 331
 FT DISULFID 122 139
 FT DISULFID 198 227
 FT DISULFID 311 383
 FT DISULFID 390 452
 FT DISULFID 403 422
 FT DISULFID 412 486
 FT DISULFID 463 507
 FT DISULFID 512 562
 FT DISULFID 532 555
 FT CARBOHYD 297 297
 FT CARBOHYD 394 394
 FT LIPID 574 574
 SQ SEQUENCE 599 AA; 63848 MW; 746730AE9E2A2E7C CRC64;
 Query Match 11.2%; Score 87; DB 1; Length 599;
 Best Local Similarity 89.5%; Pred. No. 4.4;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 108 AKNYDQVTRVVTTHMAHA 126

DB 248 ASRYDQVTRVVTTHMAHA 266

RESULT 13

GP63 LEIMA STANDARD; PET; 602 AA.
 ID -GP63 LEIMA
 AC P08148; P15906;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (promastigote surface
 DE endopeptidase).
 GN GP63.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
 RX MEDLINE=88154764; PubMed=3346625;
 RA Button L.L., McMaster W.R.;
 RT "Molecular cloning of the major surface antigen of leishmania.";
 RL J. Exp. Med. 167:724-729(1988).
 RN [2]
 RP REVISIONS.
 RA Button L.L., McMaster W.R.;
 RL J. Exp. Med. 171:589-589(1990).
 RN [3]
 RP GPI-ANCHOR.
 RX MEDLINE=91009116; PubMed=2145267;
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,
 RA Romans S.W., Bordier C.;
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
 RT the Leishmania major promastigote surface protease.";
 RL J. Biol. Chem. 265:16955-16964(1990).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=95406217; PubMed=7675788;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "Crystallization and preliminary X-ray diffraction studies of
 RT Leishmanolysin, the major surface metalloprotease from Leishmania
 RT major.";
 RL Proteins 22:58-66(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=98416698; PubMed=9739094;
 RA Schlagenhauf E., Etges R., Metcalf P.;
 RT "The crystal structure of the Leishmania major surface proteinase
 RT Leishmanolysin.";
 RL Structure 6:1035-1046(1998).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages
 CC in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
 CC -!- COFACTOR: Binds 1 zinc ion per subunit.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
 CC C14:0, C16:0, AND C18:0).
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y00647; CAA68673.1; -.
 DR PIR; P0221; P0221.
 DR PDB; 1LML; 17-SEP-97.
 DR MEROPS; M08.001; -.

| | | | |
|---|--|---|---------|
| FT | TURN | 365 | 366 |
| FT | STRAND | 369 | 370 |
| FT | HELIX | 372 | 374 |
| FT | TURN | 380 | 383 |
| FT | HELIX | 386 | 390 |
| FT | STRAND | 394 | 395 |
| FT | TURN | 396 | 397 |
| FT | STRAND | 398 | 399 |
| FT | TURN | 402 | 404 |
| FT | STRAND | 413 | 414 |
| FT | TURN | 417 | 418 |
| FT | STRAND | 421 | 425 |
| FT | STRAND | 428 | 429 |
| FT | HELIX | 435 | 437 |
| FT | TURN | 443 | 444 |
| FT | STRAND | 445 | 446 |
| FT | TURN | 450 | 454 |
| FT | STRAND | 458 | 465 |
| FT | TURN | 466 | 467 |
| FT | HELIX | 470 | 472 |
| FT | TURN | 475 | 477 |
| FT | HELIX | 478 | 480 |
| FT | TURN | 485 | 486 |
| FT | STRAND | 487 | 494 |
| FT | STRAND | 496 | 496 |
| FT | STRAND | 506 | 516 |
| FT | TURN | 517 | 520 |
| FT | STRAND | 521 | 525 |
| FT | TURN | 527 | 528 |
| FT | STRAND | 533 | 534 |
| FT | TURN | 537 | 538 |
| FT | STRAND | 540 | 542 |
| FT | HELIX | 543 | 545 |
| FT | TURN | 546 | 546 |
| FT | STRAND | 550 | 550 |
| FT | TURN | 552 | 553 |
| FT | STRAND | 555 | 557 |
| FT | HELIX | 561 | 565 |
| FT | TURN | 566 | 567 |
| FT | HELIX | 569 | 572 |
| FT | TURN | 573 | 573 |
| SEQ | SEQUENCE | 602 AA; 63953 MW; 982EF3245D87C43E CRC64; | |
| Query Match 11.2%; Score 87; DB 1; Length 602; | | | |
| Best Local Similarity 89.5%; Pred.No. 4.5; | | | |
| Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0; | | | |
| QY | 108 AKNYDQLVTRVVTHEMAHA 126 | | |
| Db | 251 ASRYDQLVTRVVTHEMAHA 269 | | |
| RESULT 14 | | | |
| GP63 LEIME | STANDARD; | PRT; | 646 AA. |
| ID_GP63 LEIME | | | |
| AC | P43150; | | |
| DT | 01-NOV-1995 (Rel. 32, Created) | | |
| DT | 01-NOV-1995 (Rel. 32, Last sequence update) | | |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) | | |
| DE | Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease) | | |
| DE | (Major surface glycoprotein) (GP63 protein) (Promastigote surface | | |
| DE | endopeptidase). | | |
| GN | GP63-C1 | | |
| OS | Leishmania mexicana. | | |
| OC | Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Leishmania. | | |
| OX | NCBI_TaxID=5665; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=MNYC/BZ/62/M379; | | |
| RC | MEDLINE=93149206; PubMed=8426614; | | |
| RA | Medina-Acosta E., Karsen R.E., Russell D.G.; | | |
| RT | "Structurally distinct genes for the surface protease of Leishmania | | |
| RT | mexicana are developmentally regulated." | | |

RL Mol. Biochem. Parasitol. 57:31-46(1993).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the
CC amastigote forms.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64394; CAA45733.1; -.
DR PIR; S19916; S19916.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -.
DR GYCOsuiteDB; P43150; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase M8.
DR Pfam; PF01457; Peptidase M8.1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; Multigene family.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 103 646 LEISHMANOLYSIN C1.
FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 267 267 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 127 144 BY SIMILARITY.
FT DISULFID 193 232 BY SIMILARITY.
FT DISULFID 316 388 BY SIMILARITY.
FT DISULFID 395 458 BY SIMILARITY.
FT DISULFID 408 427 BY SIMILARITY.
FT DISULFID 417 492 BY SIMILARITY.
FT DISULFID 469 513 BY SIMILARITY.
FT DISULFID 518 568 BY SIMILARITY.
FT DISULFID 538 561 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 646 AA; 69054 MW; F648DDC78C10B0A CRC64;
Query Match 11.2%; Score 87; DB 1; Length 646;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 108 AKNYDQLVTRVVVTHEMAHA 126
DB 253 ASRDYDQLVTRVVVTHEMAHA 271
RESULT 15
ID OMPB_RICJA STANDARD; PRT; 1656 AA.
AC C06653;
DT 30-MAY-2000 (Rel. 39; Created)
DT 30-MAY-2000 (Rel. 39; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptid].
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB003681; BAA20138.1; -.
DR InterPro; IPR006345; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69CDD5999F CRC64;
Query Match 11.2%; Score 87; DB 1; Length 1656;
Best Local Similarity 29.6%; Pred. No. 14;
Matches 45; Conservative 14; Mismatches 55; Indels 38; Gaps 8;
QY 6 VAFAAIVVSGSALAGVFWQGGGNGHNGGSSGDPSTLSIYQYGSANAALALQSDARK 65
DB 509 VLAAGAITLDGSATI-----TGDIGNGGG-----GAALQSITLANDATK 547
QY 66 SETTITQSG-----YGNQADVGGQADNSTIELTQNGFRNATIDQWAKNYDQLVTRVVTH 121
DB 548 ---TLTGGANIISANGTINFGANGTILKST---QNNIVVDCDLAIATDQ---TGVVDA 600
QY 122 EMAHANQTASDSSWVRQVGF--GNNATANQY 151
DB 601 SSLTNAQTILTISGT-----IGIIGANNNTILGQF 628
Search completed: August 2, 2004, 14:49:31
Job time : 5.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mmc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|-------------|
| 1 | 681 | 87.9 | 152 | 2 | O33802 |
| 2 | 591.5 | 76.3 | 150 | 2 | O33802 |
| 3 | 553 | 71.4 | 149 | 2 | O7X240 |
| 4 | 508.5 | 65.6 | 152 | 16 | O8CW63 |
| 5 | 431.5 | 55.7 | 150 | 2 | O7X237 |
| 6 | 385 | 49.7 | 76 | 2 | O54069 |
| 7 | 122 | 15.7 | 29 | 2 | O9S3J5 |
| 8 | 122 | 15.7 | 502 | 16 | O8EIH4 |
| 9 | 109 | 14.1 | 139 | 16 | O8EIH3 |
| 10 | 104 | 13.4 | 7716 | 16 | O7UW28 |
| 11 | 103 | 13.3 | 362 | 16 | O89D03 |
| 12 | 101 | 13.0 | 409 | 5 | O19414 |
| 13 | 100 | 12.9 | 151 | 2 | O7X238 |
| 14 | 100 | 12.9 | 362 | 16 | O8EV84 |
| 15 | 99.5 | 12.8 | 171 | 16 | O89J13 |
| 16 | 99 | 12.8 | 1748 | 5 | O94821 |

17 98 12.6 713 5 Q9N8N5
18 96.5 12.5 151 16 Q7UCZ1
19 96.5 12.5 160 16 Q8JRU7
20 96 12.4 154 16 Q89J15
21 96 12.4 157 16 Q88HG0
22 95.5 12.3 151 2 Q7X244
23 95.5 12.3 160 16 Q8CM64
24 95.5 12.3 624 3 Q8NIV1
25 95.5 12.3 1286 2 Q841Y5
26 94.5 12.2 453 5 Q9NM68
27 94 12.1 179 2 O33801
28 93.5 12.1 453 5 Q9NGF6
29 93.5 12.1 453 5 Q9NGF7
30 93.5 12.1 552 16 P96840
31 93.5 12.1 552 16 Q7TW76
32 93.5 12.1 623 16 Q8VIY0
33 93 12.0 3501 16 Q8Y106
34 93 12.0 3552 16 Q8XSD6
35 92.5 11.9 348 13 Q93397
36 92.5 11.9 440 16 Q92DG1
37 92.5 11.9 1222 5 Q9W4F0
38 92.5 11.9 1222 5 Q8T8L9
39 92.5 11.9 1351 5 Q8SX56
40 92.5 11.9 1354 5 Q8MPN4
41 91.5 11.8 1422 16 Q8EFU3
42 91.5 11.8 1615 2 Q9KKA8
43 91 11.7 191 3 Q8TFA6
44 90.5 11.7 152 2 Q7X241
45 90.5 11.7 1209 16 Q89CK5

ALIGNMENTS

RESULT 1

O33802 ID O33802 PRELIMINARY; PRT; 152 AA.
AC O33802;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AGFA protein (Fragment).
GN AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98033981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells."
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CAA04151.1; -;
FT NON TER 152
SQ SEQUENCE 152 AA; 15401 MW; 9DAYDADC2364B006 CRC64;

Query Match 87.9%; Score 681; DB 2; Length 152;
Best Local Similarity 90.1%; Pred. No. 2e-46;
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSAVAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVCGGADNSITLTQNGFRNATIDQNAKYNQLYTRVVT 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSITLTQNGFRNATIDQNAKNSDITVQGYG 120
QY 121 HEMAHANQTASDSSVMVRQVFGNNATANQY 151

Best Local Similarity 29.1%; Pred. No. 81;
Matches 39; Conservative 18; Mismatches 57; Indels 20; Gaps 6;
QY 28 GCGNHN--GGGNS--GPDSTLSIYQGSANAALQSDARKSETTITQSGY 75
Db 4048 GCGNHTLGGVAVSQPTSSSVGEGTSLV---SVGNGLANDIEMDGDLSITTEIN- 4103
QY 76 GNGADYVQCG---ADNSTIILTQNGFENNATIDQWNAKNYDQLVTRVVTHEMAHQAOTASD 132
Db 4104 GNSGVGNITLPGATIQLEADGSSYDPGTIYQNLNDGATATERTFTVSDGNGTDT 4163
QY 133 SSVMVRQVGFNNA 146
Db 4164 TSAITIIIG-GNDA 4176
RESULT 11
Q89D03 PRELIMINARY; PRT; 362 AA.
AC Q89D03;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical exported glutamine-rich protein.
GN BLU7642.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110.
RX MEDLINE=244998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005962; BAC52907.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 362 AA; 39058 MW; D447EED04E8433EC CRC64;
Query Match 13.3%; Score 103; DB 16; Length 362;
Best Local Similarity 31.6%; Pred. No. 26;
Matches 50; Conservative 14; Mismatches 80; Indels 14; Gaps 6;
QY 2 KLLKVAFAAIVVSGSALAGVFPQWGGGNGHNGSSGSDST----LSIYQYGSANAAL 57
Db 3 KRLFLATTAATAVATSAQAQSSP---STSNPNSTTQRPDSTSTTPSSSTPSSQAQTNP 59
QY 58 ALQSDARKSETTITQSGYNGADVQGGADNSTIILTQNGFENNATI----DOWNAKNYDQ 113
Db 60 STNSAQTSPPSTGQSAAGQTTSCT-NTTAQTSNNSTNQATQSQPSQNTNAPS-DQ 117
QY 114 LVTRVVTHEMAH-ANQTPASDSVMVRQVGFNNAQ 150
Db 116 TQTPPTDNRQAQSANPPASGASQAQSPGSGNSTNTAQ 155
RESULT 12
Q19414 PRELIMINARY; PRT; 409 AA.
AC Q19414;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE F13E9.4 protein.
GN F13E9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN S00866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Bean M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Uppraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A., C.M.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AS015532; AAN53942.1; -
DR TIGR; S00866; -
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41ECLCPA76957920 CRC64;
Query Match 14.1%; Score 109; DB 16; Length 139;
Best Local Similarity 30.1%; Pred. No. 0.28;
Matches 34; Conservative 19; Mismatches 46; Indels 14; Gaps 3;
QY 39 SGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVQGGADNSTIILTQNGFR 98
Db 41 SGRDNLIDLQQQTANQGVFQSGSDNS-AYYTAGNDNISLVTQGTGNNVQLLVQGAQ 99
QY 99 NNATIDQWNAKNYDQLVTRVVTHEMAHQAOTASDSVMVRQVGFNNAQ 151
Db 100 NKASITQIGNDNLVQL-----NQLGS-GNFSIQIADGAISITQY 139
RESULT 10
Q7UWZ8 PRELIMINARY; PRT; 7716 AA.
AC Q7UWZ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB1661.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetiales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Glickner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294135; CAB72214.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 7716 AA; 797868 MW; D391A25BD96405C0 CRC64;
Query Match 13.4%; Score 104; DB 16; Length 7716;

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
none;
RA "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z69383; CAA93412.1; -.
DR FIR: T20847; T20847.
DR WormPep: F13E9.4; CE05606.
DR InterPro: IPR003677; Onchoerca_Ag.
DR Pfan; PF02520; DUF148; 1.
SQ SEQUENCE 409 AA; 43231 MW; B07DFOE4175C5739 CRC64;

Query Match 13.0%; Score 101; DB 5; Length 409;
Best Local Similarity 23.3%; Pred. No. 4.3;
Matches 37; Conservative 26; Mismatches 65; Indels 30; Gaps 5;

QY 15 SGALAGVVPWGGGNGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSG 74
DB 134 SGQAGSMNFGGGGQYQNGQNGFGGQSGGSGNSLSANSNGNNQG--SSSG 191

QY 75 YGN--GADVGGGANSTIE---LTQNGFRNN-----ATIDQWNAK 109
DB 192 YQNNQGRHQGGGSHSSNSVMSNGYSSNGYNNNGPTPSPFLNNVSSSAQDYTNIV 251

QY 110 NYDQLVTRVVTHEMAHANQTDSDSVVMVQVQFGGNATA 148
DB 252 NKSLLTNQINEQ---ASNWASNSVQAQYIQYETNRSA 287

RESULT 13
Q7X238
ID Q7X238 PRELIMINARY; PRT; 151 AA.
AC Q7X238;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE "Production of Cellulose and Curli Fimbriae by Members of the Family
DE Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
DE Nucleation component of curlin monomers.
GN CSGB.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56677.1; -.
SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A27882B7 CRC64;

Query Match 12.9%; Score 100; DB 2; Length 151;
Best Local Similarity 31.8%; Pred. No. 1.6;
Matches 35; Conservative 12; Mismatches 45; Indels 18; Gaps 4;

QY 10 AAVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALALQSDARKSETT 69
DB 58 AQIRQSGKLLSVVQ-----DGAGNRAEVD-----QSGTYNIADIQS--GNGNDAG 103

QY 70 ITQSGYGNAGVGGGADNSITELTQNGFRNNATTDQWNAKNDQLVTRVV 119
DB 104 ITQDGYGNSAKIIQKSGNRRANITQGTQKTAVVVQ-----KQSQMAIRVI 149

RESULT 14

```

```

Q8EV84
ID Q8EV84 PRELIMINARY; PRT; 362 AA.
AC Q8EV84;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P35 lipoprotein homolog.
DE MYPE6840.
GN MYCOPLASMA PENETRANS.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=26227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=12466555;
RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300(2002).
DR EMBL: AF004172; BAC44476.1; -.
DR InterPro: IPR000437; PROKAR_LIPOPROTEIN; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 362 AA; 39547 MW; 8DA27F70D19D354F CRC64;

Query Match 12.9%; Score 100; DB 16; Length 362;
Best Local Similarity 23.9%; Pred. No. 4.5;
Matches 42; Conservative 33; Mismatches 63; Indels 38; Gaps 8;

QY 1 MKLLKVAAPAAIVVSGS-ALAGVVP-----QWGGGNGHNGGNSG-----PDS 43
DB 1 MKIKKILKALALTGAFIVATVPVIVSSCSTSDNNGNGNNNGNQDGGGQOOTEI 60

QY 44 TLSIYQGSANAALALQSDARKSETTITQSG-----YNGADYGGQADNSTIEL--- 92
DB 61 TPTIKKEVSLSGALSKIYDANKSTSLIAEDIKANPTNYDFNGEALKDLIKDATVSVNG 120

QY 93 -TQNGFRNNATTDQWNAKNDQLVTRVVTHEMAHANQTDSDSV--MYRQVFGNN 145
DB 121 FTSTFKGD-TVETWSAKYGVKKGT-----YQAASKQLDIKSINDLETQLGDSNN 169

RESULT 15
Q89JI3
ID Q89JI3 PRELIMINARY; PRT; 171 AA.
AC Q89JI3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CsgA protein.
DE CSGA OR BLJ5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL: AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 12.8%; Score 99.5; DB 16; Length 171;
Best Local Similarity 25.8%; Pred. No. 2;

```

| Matches | 40; | Conservative | 21; | Mismatches | 63; | Indels | 31; | Gaps | 4; |
|---------|-----|--|-----|------------|-----|--------|-----|------|----|
| QY | 1 | MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ | 60 | | | | | | |
| Db | 40 | MEKLPFASVAVLALSSAAQA-----NTSTTVQGLVNGSSVTQ | 78 | | | | | | |
| QY | 61 | SDARKSETTITQSGYNGADVQGGAD---NSTIELTQNGFRNNATIDQWNAKNDQLVT | 116 | | | | | | |
| Db | 73 | NGLTNDSSSTTQIGILNGASTMQGTSSPFLNNVSTVNCAGVQNSATTGQVAFGNNGSAIT | 138 | | | | | | |
| QY | 117 | RVVTHEMAHANOTASDSSVMVRQVGF--NNATANQ | 150 | | | | | | |
| Db | 139 | QNSFGPPALQNNAS-----VGQLSFGINTSTVSQ | 168 | | | | | | |

Search completed: August 2, 2004, 14:54:40
Job time : 29.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_29Jan04.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 782 | 100.0 | 151 | 3 | AAB36353 | Aab36353 Agfa::PT3 |
| 2 | 709 | 90.7 | 151 | 3 | AAB36350 | Aab36350 Agfa::PT3 |
| 3 | 692 | 88.5 | 151 | 2 | AAR74625 | Aar74625 Agfa sequ |
| 4 | 692 | 88.5 | 151 | 3 | AAB36341 | Aab36341 Salmonell |
| 5 | 687 | 87.9 | 151 | 2 | AAW23370 | Aaw23370 Salmonell |
| 6 | 675 | 86.3 | 151 | 3 | AAB36349 | Aab36349 Agfa::PT3 |
| 7 | 617 | 78.9 | 151 | 3 | AAB36354 | Aab36354 Agfa::PT3 |
| 8 | 614 | 78.5 | 151 | 3 | AAB36346 | Aab36346 Agfa::PT3 |
| 9 | 612 | 78.3 | 151 | 3 | AAB36347 | Aab36347 Agfa::PT3 |
| 10 | 609 | 77.9 | 151 | 3 | AAB36352 | Aab36352 Agfa::PT3 |
| 11 | 601 | 76.9 | 151 | 3 | AAB36351 | Aab36351 Agfa::PT3 |
| 12 | 600 | 76.7 | 151 | 3 | AAB36355 | Aab36355 Agfa::PT3 |
| 13 | 577 | 73.8 | 151 | 3 | AAB36348 | Aab36348 Agfa::PT3 |
| 14 | 523 | 66.9 | 151 | 3 | AAB36343 | Aab36343 Escherich |
| 15 | 518 | 66.2 | 151 | 7 | ABR82651 | ABr82651 E. coli C |
| 16 | 507 | 64.8 | 120 | 2 | AAR62761 | Aar62761 Agfa sequ |
| 17 | 507 | 64.8 | 120 | 2 | AAW23359 | Aaw23359 Salmonell |
| 18 | 445 | 56.9 | 142 | 2 | AAK52864 | Aar52864 Fibronect |
| 19 | 373 | 47.7 | 122 | 2 | AAR52863 | Aar52863 FNB curli |
| 20 | 188 | 24.0 | 45 | 3 | AAB36316 | Aab36316 Salmonell |
| 21 | 132 | 16.9 | 22 | 3 | AAB36318 | Aab36318 Salmonell |
| 22 | 123 | 15.7 | 23 | 3 | AAB36321 | Aab36321 Salmonell |
| 23 | 123 | 15.7 | 23 | 3 | AAB36326 | Aab36326 Salmonell |
| 24 | 123 | 15.7 | 23 | 3 | AAB36338 | Aab36338 Salmonell |
| 25 | 113 | 14.5 | 24 | 7 | ABR82844 | ABr82844 E. coli c |

| | | | | | | |
|----|------|------|------|---|----------|--------------------|
| 26 | 112 | 14.3 | 151 | 3 | AAB36344 | Aab36344 Escherich |
| 27 | 111 | 14.2 | 22 | 3 | AAB36322 | Aab36322 Salmonell |
| 28 | 111 | 14.2 | 22 | 3 | AAB36327 | Aab36327 Salmonell |
| 29 | 111 | 14.2 | 22 | 3 | AAB36337 | Aab36337 Salmonell |
| 30 | 107 | 13.7 | 151 | 3 | AAB36342 | Aab36342 Salmonell |
| 31 | 105 | 13.4 | 597 | 4 | AAU08231 | Aau08231 Polypepti |
| 32 | 102 | 13.0 | 26 | 7 | ABR82649 | ABr82649 E. coli V |
| 33 | 100 | 12.8 | 943 | 2 | AAW64378 | Aaw64378 Mycobacte |
| 34 | 100 | 12.8 | 943 | 2 | AAW61745 | Aaw61745 M. tuberc |
| 35 | 100 | 12.8 | 943 | 2 | AAV39032 | Aay39032 M. tuberc |
| 36 | 100 | 12.8 | 943 | 2 | AAV39175 | Aay39175 M. tuberc |
| 37 | 100 | 12.8 | 3300 | 6 | ABU36445 | Abu36445 Protein e |
| 38 | 98.5 | 12.6 | 738 | 2 | AAW56163 | Aaw56163 New DNA s |
| 39 | 97.5 | 12.5 | 850 | 4 | ABB65764 | Abb65764 Drosophil |
| 40 | 97.5 | 12.5 | 1028 | 4 | ABB62708 | Abb62708 Drosophil |
| 41 | 96.5 | 12.3 | 447 | 3 | AAG29728 | Aag29728 Arabidops |
| 42 | 96.5 | 12.3 | 468 | 3 | AAG29727 | Aag29727 Arabidops |
| 43 | 96 | 12.3 | 19 | 3 | AAB36323 | Aab36323 Salmonell |
| 44 | 96 | 12.3 | 19 | 3 | AAB36336 | Aab36336 Salmonell |
| 45 | 96 | 12.3 | 19 | 3 | AAB36328 | Aab36328 Salmonell |

ALIGNMENTS

RESULT 1
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
AC AAB36353;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN W0200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/PAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 782; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.3e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVGQYG 120
 DB 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVGQYG 120
 QY 121 NNALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2
 AAB36350
 ID AAB36350 standard; protein; 151 AA.
 XX
 AC AAB36350;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::P3#5 amino acid sequence SEQ ID NO:20.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 FN WO2000060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI: 2000-672631/65.
 DR N-PSDB; AAC64626.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEFI7/TAFF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX

SQ Sequence 151 AA;

Query Match 90.7%; Score 709; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 2.5e-60;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAAL--- 57
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALQ 60
 QY 58 -----YDQLVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSD 112
 DB 61 SDARKYDQLVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSD 112
 QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 XX
 AC AAR74625;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine.
 XX
 OS Salmonella.
 XX
 FN WO9425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 XX
 XX 26-APR-1993; 93US-00054452.
 XX
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI; 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct FN field.)
 XX
 SQ Sequence 151 AA;
 Query Match 88.5%; Score 692; DB 2; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 DT 26-FEB-2001 (first entry)
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 PN WO2000060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collinson SK, Kay WW;
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64617.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 88.5%; Score 692; DB 3; Length 151;
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGDPSTLSIYQYGSANAALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 AC AAW23570;
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa.
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;

```

XX WPI; 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfA encoded by the full agfA gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteriaceae. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridize to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
SQ
Query Match      87.9%; Score 687; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 3.3e-58;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L Y D Q 60
D b 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L A L Q 60
QY 61 L V T R V V T H E M A H A G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Y G G 120
D b 61 S D A R K S E T T I T Q G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Y G G 120
QY 121 N N A A L V N Q T A S D S S V V M V R Q V G F G N N A T A N Q Y 151
D b 121 N N P A L V N Q T A S D S S V V M V R Q V G F G N N A T A N Q Y 151
RESULT 6
AAB36349
ID AAB36349 standard; protein; 151 AA.
AC AAB36349;
XX
XX 26-FEB-2001 (first entry)
XX
XX AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant AgfA

```

```

PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal, AgfA
XX comprising separating an amino acid polymer comprising a recombinant amino
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant AgfA protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match      86.3%; Score 675; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-57;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L Y D Q 60
D b 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L Y D Q 45
QY 61 L V T R V V T H E M A H A G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q 105
D b 46 L V T R V V T H E M A H A L Q S D A R K S E T T I T Q G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q 105
QY 106 W N A K N S D I T V G Y G G N N A A L V N Q T A S D S S V V M V R Q V G F G N N A T A N Q Y 151
D b 106 W N A K N S D I T V G Y G G N N A A L V N Q T A S D S S V V M V R Q V G F G N N A T A N Q Y 151

```

```

RESULT 7
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
XX AAB36354;
XX
XX 26-FEB-2001 (first entry)
XX
XX AgfA::PT3#9 amino acid sequence SEQ ID NO:28.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX
XX

```


AAB36347
 ID AAB36347 standard; protein; 151 AA.
 AC AAB36347;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 N-PSDB; AAC64623.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.3%; Score 612; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-51;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60

QY 61 LVTRVVTHEMAHAGYGNADVGGADNSTIELTQNGFRNATIDOWNAKNSDITVGQYGG 120
 DB 61 SDARKSETTITQSGYGNADVGGADNSTIELTQNGFRNATIDOWNAKNSDITVGQYDQ 120
 QY 121 NNAALVNOTASDSSVMVRQVGFQGNATANY 151
 DB 121 LVTRVVTHEMAHASVMVRQVGFQGNATANY 151
 RESULT 10
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 AC AAB36352;
 XX
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 N-PSDB; AAC64628.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.3%; Score 612; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 5.5e-51;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQGSANAALYDQ 60

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 77.9%; Score 609; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 1.1e-50;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNDYQLVTRVVT 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 HMAHANGTASDSSVMVRQVGFNNATANQY 151

CC Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
SQ useful for the expression of recombinant Agfa protein which is useful for
eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
SQ Sequence 151 AA;

Query Match 76.9%; Score 601; DB 3; Length 151;
Best Local Similarity 81.5%; Pred. No. 6.3e-50;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
AAB36351
ID AAB36351 standard; protein; 151 AA.
XX
AC AAB36351;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
DN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
FA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
XX
DR N-PSDB; AAC64627.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

RESULT 12
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
FA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
XX
DR N-PSDB; AAC64631.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 76.7%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. NO. 7.8e-50;
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 LVTRVVTTHMAHAGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETLTQNGFRNNATYDQLVTRVVTTHMAHAGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13

AAB36348
 ID AAB36348 standard; protein; 151 AA.
 XX AAB36348;
 XX 26-FEB-2001 (first entry)
 XX AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO2000060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JU, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.8%; Score 577; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.3e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHAGSPPDSTLSIYQYGSANAALALQ 60
 QY 61 LVTRVVTTHMAHAGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETLTQNGFRNNATIDOWNAKNSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14

AAB36343
 ID AAB36343 standard; protein; 151 AA.
 XX AAB36343;
 XX 26-FEB-2001 (first entry)
 XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 XX Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen.
 XX Escherichia coli.
 XX WO2000060102-A2.
 XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888F.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI; 2000-672631/65.
 DR N-PSDB; AAC64619.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEf1/TAf) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 66.9%; Score 523; DB 3; Length 151;
 Best Local Similarity 68.9%; Pred. No. 2e-42;
 Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGSPDSTLSIYQYGSANAALYDQ 60
 Db 1 MKLLKVAFAAIVFSGSALAGVVPQYGGGNGHGGNSGSPNSLNIIYQYGGNSALALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVCGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 Db 61 TDARNSDLTITQHGGGNGADVCGQSDSSIDLTRQFGNSATLDQWNGKNSMTYKQFGG 120
 QY 121 NNAALVNQASDSSVMVROVGFNNATANQY 151
 Db 121 GNGRAVDQTASNSVNVTVQVGFNNATAHQY 151
 RESULT 15
 ABR82651
 ID ABR82651 standard; protein; 151 AA.
 XX
 AC ABR82651;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE E. coli CsgA subunit 15 kDa protein.
 XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
 XX
 OS Escherichia coli.
 XX
 PN WO2003064446-A2.
 XX
 PD 07-AUG-2003.
 XX
 XX 30-JAN-2003; 2003WO-BP000943.
 XX
 XX 31-JAN-2002; 2002GB-00002275.
 PR
 PA (HANS-) HANSA MEDICAL RES AB.
 XX
 PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;
 XX
 DR WPI; 2003-646136/61.
 DR N-PSDB; ACR36153.
 XX
 XX New isolated peptide capable of binding a mammalian plasma protein,
 PT useful in the manufacture of a medicament for the prevention and/or
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
 PT or Shigella infections.
 XX
 PS Disclosure; Page 41-42; 42pp; English.
 XX
 CC The invention relates to an isolated peptide capable of binding a
 CC mammalian plasma protein or of generating an immune response in a mammal
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
 CC antibody is useful for treating a bacterial infection in a human or
 CC animal or in the manufacture of a medicament for the prophylactic
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
 CC or Shigella infection. The peptide that is immobilized on a solid support
 CC is also useful as a reagent for determining the ability of a plasma
 CC protein to bind to bacteria. The present sequence represents an E. coli
 CC 15 kDa protein
 XX
 SQ Sequence 151 AA;
 Query Match 66.2%; Score 518; DB 7; Length 151;
 Best Local Similarity 68.2%; Pred. No. 6.2e-42;
 Matches 103; Conservative 18; Mismatches 30; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGSPDSTLSIYQYGSANAALYDQ 60
 Db 1 MKLLKVAFAAIVFSGSALAGVVPQYGGGNGHGGNSGSPNSLNIIYQYGGNSALALQ 60
 QY 61 LVTRVVTHEMAHAGYNGADVCGQADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
 Db 61 TDARNSDLTITQHGGGNGADVCGQSDSSIDLTRQFGNSATLDQWNGKNSMTYKQFGG 120
 QY 121 NNAALVNQASDSSVMVROVGFNNATANQY 151
 Db 121 GNGRAVDQTASNSVNVTVQVGFNNATAHQY 151
 Search completed: August 2, 2004, 14:48:28
 Job time : 45.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIIWVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued Patents AA:*
- 2: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 687 | 87.9 | 151 | 1 | US-08-233-788A-59 |
| 2 | 507 | 64.8 | 120 | 1 | Sequence 59, Appl |
| 3 | 100 | 12.8 | 943 | 4 | US-08-233-788A-57 |
| 4 | 100 | 12.8 | 943 | 4 | Sequence 57, Appl |
| 5 | 100 | 12.8 | 943 | 4 | Sequence 204, App |
| 6 | 100 | 12.8 | 943 | 4 | Sequence 199, App |
| 7 | 98.5 | 12.6 | 738 | 3 | Sequence 131, App |
| 8 | 91.5 | 11.9 | 892 | 4 | Sequence 204, App |
| 9 | 86 | 11.3 | 273 | 4 | Sequence 3, Appl |
| 10 | 82 | 10.5 | 673 | 3 | Sequence 5, Appl |
| 11 | 82 | 10.5 | 673 | 3 | Sequence 4764, Ap |
| 12 | 82 | 10.5 | 673 | 3 | Sequence 6167, Ap |
| 13 | 82 | 10.5 | 673 | 3 | Sequence 26438, A |
| 14 | 82 | 10.5 | 673 | 3 | Sequence 8, Appl |
| 15 | 82 | 10.5 | 673 | 3 | Sequence 8, Appl |
| 16 | 82 | 10.5 | 673 | 3 | Sequence 10, Appl |
| 17 | 82 | 10.5 | 673 | 3 | Sequence 10, Appl |
| 18 | 82 | 10.5 | 673 | 3 | Sequence 2, Appl |
| 19 | 81.5 | 10.4 | 873 | 4 | Sequence 2, Appl |
| 20 | 81 | 10.4 | 2736 | 4 | Sequence 13, Appl |
| 21 | 80 | 10.2 | 863 | 4 | Sequence 30227, A |
| 22 | 79 | 10.1 | 353 | 2 | Sequence 10179, A |
| 23 | 78.5 | 10.0 | 745 | 4 | Sequence 3, Appl |
| 24 | 78.5 | 10.0 | 1739 | 4 | Sequence 6, Appl |
| 25 | 78 | 10.0 | 1216 | 4 | Sequence 3739, Ap |
| 26 | 77.5 | 9.9 | 321 | 4 | Sequence 5130, Ap |
| 27 | 77.5 | 9.9 | 713 | 4 | Sequence 18, Appl |
| | | | | | Sequence 53, Appl |

| | | | | | | |
|----|------|-----|------|---|---------------------|-------------------|
| 28 | 77 | 9.8 | 1338 | 2 | US-08-728-470-9 | Sequence 9, Appl |
| 29 | 77 | 9.8 | 1338 | 3 | US-08-719-641-9 | Sequence 9, Appl |
| 30 | 77 | 9.8 | 1529 | 2 | US-08-728-470-10 | Sequence 10, Appl |
| 31 | 77 | 9.8 | 1529 | 3 | US-08-719-641-10 | Sequence 10, Appl |
| 32 | 77 | 9.8 | 1599 | 2 | US-08-617-697-9 | Sequence 9, Appl |
| 33 | 77 | 9.8 | 1600 | 2 | US-08-617-697-10 | Sequence 10, Appl |
| 34 | 77 | 9.8 | 2315 | 4 | US-08-543-681A-5434 | Sequence 5434, Ap |
| 35 | 76.5 | 9.8 | 363 | 1 | US-08-458-023B-6 | Sequence 6, Appl |
| 36 | 76.5 | 9.8 | 435 | 2 | US-08-331-515A-2 | Sequence 2, Appl |
| 37 | 76.5 | 9.8 | 435 | 3 | US-09-168-406A-2 | Sequence 2, Appl |
| 38 | 76.5 | 9.8 | 941 | 4 | US-09-336-447A-9 | Sequence 9, Appl |
| 39 | 76.5 | 9.8 | 1612 | 1 | US-08-169-927-2 | Sequence 2, Appl |
| 40 | 76 | 9.7 | 364 | 1 | US-07-792-259-17 | Sequence 17, Appl |
| 41 | 76 | 9.7 | 1385 | 1 | US-07-876-280-2 | Sequence 2, Appl |
| 42 | 76 | 9.7 | 1385 | 1 | US-07-675-772-2 | Sequence 2, Appl |
| 43 | 76 | 9.7 | 1385 | 1 | US-08-063-170-2 | Sequence 2, Appl |
| 44 | 76 | 9.7 | 1385 | 1 | US-08-158-232-2 | Sequence 2, Appl |
| 45 | 76 | 9.7 | 1385 | 1 | US-08-304-626-2 | Sequence 2, Appl |

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouchier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEDDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-788A-59

Query Match 87.9%; Score 687; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 78-60;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIIWVSGSALAGVVPQGGGNNHGGNSGPDSTLSIYQYGSANAALYDQ 60
|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALALQ 60
 QY 61 LVTRVVVTHEMAHAGYNGADVCGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
 Db 61 SPARKSETTITSGYNGADVCGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVNVQVGGNNATANQY 151
 Db 121 NNPALVNOTASDSSVNVQVGGNNATANQY 151

RESULT 2

US-08-233-788A-57
 ; Sequence 57, Application US/08233788A
 ; Patent No. 5635617
 ; GENERAL INFORMATION:
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Kay, William W.
 ; APPLICANT: Collinson, Karen S.
 ; APPLICANT: Clouchier, Sharon C.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 ; OF SALMONELLA
 ; NUMBER OF SEQUENCES: 61
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Seed and Berry
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: U.S.A.
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/233,788A
 ; FILING DATE: 26-APR-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: King, Joshua
 ; REGISTRATION NUMBER: 35,570
 ; REFERENCE/DOCKET NUMBER: 920043.403C2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; TELEX: 3723836 SEDANBERRY
 ; INFORMATION FOR SEQ ID NO: 57:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 120 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-233-788A-57

Query Match 64.8%; Score 507; DB 1; Length 120;
 Best Local Similarity 87.5%; Pred. No. 2.1e-42;
 Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 QY 22 VVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALYDQVTRVTHEMAHAGYNGADV 81
 Db 1 VVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSYGNGADV 60
 QY 82 QGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGNNAALVNOTASDS 133
 Db 61 QGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGNNAALVNOTASDS 112

RESULT 3

US-09-056-556-204
 ; Sequence 204, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/056,556
 FILING DATE: 07-APR-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.457
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 204:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 943 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-09-056-556-204

Query Match 12.8%; Score 100; DB 4; Length 943;
 Best Local Similarity 26.0%; Pred. No. 0.2;
 Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;
 QY 16 GSALAGVVPQWGGG-GNHN-GCGNSSGPDSTLSIYQYGSANAALYDQVTRVTHEMAHA 73
 Db 464 GSCNIGVFNVGSGSLGNYNIGSN-----LGIYNIQVGNVDY-----NV 503
 QY 74 GYNGADVCGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGN-----121
 Db 504 GFENAGDFNQGFANT-----GNNNIGFANTGNNNIGLGLSGDNNQGFNIAGWNS 553
 QY 122 ---NNAALVNOTASDSSVM---VRQVGFNNATAN 149
 Db 554 GTGNSGLFNSGTTNVGIFNAGTGNVGIANSGTGN 587

RESULT 4

US-09-072-596-199
 ; Sequence 199, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonia
 APPLICANT: Houghton, Raymond
 APPLICANT: Houghton, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 NUMBER OF SEQUENCES: 350
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle

TUBERCULOSIS

```
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-596-199

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNVGDY-----NV 503

QY 74 GYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYCGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNIASGWS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
Db 554 GTGNSGLFNSGTNNVGFNAGTGNVGIANSGTGN 587

RESULT 5
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNVGDY-----NV 503

QY 74 GYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYCGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNIASGWS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
Db 554 GTGNSGLFNSGTNNVGFNAGTGNVGIANSGTGN 587

RESULT 6
US-09-072-967-204
; Sequence 204, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-204

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNVGDY-----NV 503

QY 74 GYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYCGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNIASGWS 553
```

QY 122 ---NAALVNQTASDSSVM---VRQVGFGNATAN 149
 : | | : : | | | : | |
Db 554 GTGNSGLFNCGTNNVGFNAGTCNVGIANSGTGN 587

RESULT 7
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR. POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

Query Match 12.6%; Score 98.5; DB 3; Length 738;
Best Local Similarity 28.1%; Pred. No. 0.2;
Matches 45; Conservative 11; Mismatches 55; Indels 49; Gaps 6;

Qy 3 LLKVAFAAIVVSGSALAGWPQWGGGNGHGGNSSGGPDSTLSTYQYGSANAALYDQLV 62
||||| : |||||
Db 419 LLIKSSASASASASASASAG-----GGGGGNGGNGGGGG-----GGAGA----- 458

Qy 63 TRWVTHEAHAGYNGGADVGGADNSTIELTQ-----NGFRNNATIDQWNAKNSDITVGQ 117
: : | | | | | : | | | :
Db 459 ---LAALAAAGAGGGGCGGGGGGALAAALAAAGAGGGGFGGLGGL-----GG 503

Qy 118 YGNNAAALVNQTASDSS-----VMVRQVGFNNATA 148
Db 504 LGGGSAIAAAAAAAAAASGGGPRALRRALRRQMRGGGSAIA 543

RESULT 8
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190

| | | | | |
|-----------------------|--------|-----------------|-------|----------------|
| Query Match | 11.9%; | Score 93; | DB 4; | Length 892; |
| Best Local Similarity | 27.7%; | Pred. No. 0.89; | | |
| Matches | 41; | Conservative | 14; | Mismatches 49; |
| | | | | Indels 44; |
| | | | | Gaps 9; |

QY 28 GGGNHN-----GGGNSS---GPDSTLSIYQGSAANAALYDQLVTRVVTHEMAFAGY 75
|||::| |||::| |||::| |||::|
Db 89 GGGDYNEAKGNYSTVGGSNSTAKGEKSTIGGGDTNDANG-----TYSITGGGY 137

QY 76 ----GNGADVGGADNSTI---ELTQGFGRNATIDQNAKNSDITVGQYG---GNAAL 125
 | | | | | | | | | | | | | | | |
Db 130 YSRAIGDSSTICGGYNQATGEKSTVACGRNN----QATGNNSIVAGGSVNQATGNNSIV 193

Qy 126 V---NQ TASDSSVMVRQVGFGNNATAN 149
 ||| : |||
Db 194 AGGSHNQATGEGSF---AAGVENKANAN 218

RESULT 9
US-09-328-352-4764
; Sequence 4764, Application US/09328352

Query Match 11.7%; Score 91.5; DB 4; Length 975;
Best Local Similarity 27.4%; Pred. No. 1.4;
Matches 37; Conservative 19; Mismatches 68; Indels 11; Gaps 4;

[illegible]

Qy 75 ---YCGADYGQGDNSTIELTQGFRNNATIDQWAKNSDITVGQYGGNNAALVNQT 129
 | : | : | : | : | : | : | :
Dd 338 NSFTLIGNSS---SSSWNTAPTTSVTNDNTID--NGNSGGTGSGGNGSGDGILLNGA 392

QY 130 ASDSSVMVRQVGFN 144
|| : ||

Db 393 ASNGEHNHYGIGNGN 407

RESULT 10

US-09-328-352-6167

Sequence 6167, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 6167

LENGTH: 273

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-6167

Query Match 11.3%; Score 88; DB 4; Length 273;

Best Local Similarity 28.5%; Pred. No. 0.6;

Matches 43; Conservative 19; Mismatches 65; Indels 24; Gaps 7;

Qy 1 MKLKVA--PAALIVSGSALAGVVPWGGGNNH-----NGGNS--GPDSTLSIYQYQSA 53

Db 19 MKLIAASALLSALAVGAANA-YQAEVGGSYNLDPDNGSSVSKFGVGTGYFNPVQTR 77

Qy 54 NAALYDQIVTRVVTHEMAHAGYNGADVGGQADNSTIELTQNGFRNNATIDOWNAKNSDI 113

Db 78 NAPLAERAFNLNRASNVNHNHYG-----DMSGTRDTQYGVG-----VEYFVPSDF 123

Qy 114 TVGOYGGNNAALVNQTSDDSSV--MVRQVGF 142

Db 124 YLSGDVGRNEREIDNTWIDSKVTYYAAEVGY 154

RESULT 11

US-09-252-991A-26438

Sequence 26438, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 26438

LENGTH: 1415

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26438

Query Match 11.0%; Score 86; DB 4; Length 1415;

Best Local Similarity 23.9%; Pred. No. 7.8;

Matches 38; Conservative 21; Mismatches 68; Indels 32; Gaps 5;

Qy 13 VVSGSALAGVVPWGGGNNH-----GNSSGPDSSTLSIYQYGSANAALYDQIVT 63

Db 366 ILSHEVSVAAVQANAGSGQVHVAGPGAGANSNGVTIVQQPAPVDLAAGANGTS 425

Qy 64 RVWTHEMAHAGY-----NGADYGGQADNSTIELTON-----GFRNNATIDOWN 107

Db 426 AVQSOSGANIGSANGISVWSQGANIGAGASDISVVQSQNSNIGSGYNGVTIVQSQN 485

Qy 108 AKN-----SDITVGOYGGNNAALVNQTSADSSVMVRQVG 141

Db 486 GANIGSGASGITVWQ--SQNGANIGSGASGISVWQSQSG 522

RESULT 12

US-09-196-387-8

Sequence 8, Application US/09196387

Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-196-387-8

Query Match 10.5%; Score 82; DB 3; Length 673;

Best Local Similarity 29.6%; Pred. No. 7.4;

Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVSGSALAGVVPWGGGNNHNGGNSGPDSTLSIYQYGSANAALYDQIVT 61

Db 99 VAAAPVVPVAVTSSAAGVAPNAGSGNNPSSSSSPTSS--SSSSPSPGSSLAESPEAA 157

Qy 62 -VTRVVTHEMAHAGYNGADVGGQADNSTIELTQNG--FRNNATIDOWNAKNSDI 113

Db 158 GVSSTAPLGFGAAGPGCTGVPAVSGALRELLEACRNGDVSRVRLVDAAVNAKDM 212

RESULT 13

US-09-841-835-8

Sequence 8, Application US/09841835

Patent No. 6536587

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

```
; ADDRESS: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841.835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196.387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-841-835-8

Query Match 10.5%; Score 82; DB 4; Length 673;
Best Local Similarity 29.6%; Pred. No. 7.4;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSGGPDSTLSIYQVGSANAALYDQI--- 61
DB 99 VAAFPVPAVSTSSAAGVAPNPAAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 62 -VTRVVTHEMAHAGYNGADVQGGADNSTIELTQNG--FRNNATIDQWNAKNSDI 113
DB 158 GVSSTAPLPGNAGPGTGVPVAVSGALRELLACRNGDYSRVKRLVDAANVNAKDM 212

RESULT 14
US-09-196-387-10
; Sequence 10, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196.387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196.387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; MOLECULE TYPE: protein
; US-09-196-387-10

Query Match 10.5%; Score 82; DB 3; Length 949;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSGGPDSTLSIYQVGSANAALYDQI--- 61
DB 99 VAAFPVPAVSTSSAAGVAPNPAAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 62 -VTRVVTHEMAHAGYNGADVQGGADNSTIELTQNG--FRNNATIDQWNAKNSDI 113
DB 158 GVSSTAPLPGNAGPGTGVPVAVSGALRELLACRNGDYSRVKRLVDAANVNAKDM 212

RESULT 15
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841.835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196.387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; MOLECULE TYPE: protein
; US-09-196-387-10
```

Search completed: August 2, 2004, 14:58:36
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)
1287.123 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/FCCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 520 | 66.5 | 151 | 12 | US-09-741-873B-4 |
| 2 | 520 | 66.5 | 151 | 12 | US-09-741-873B-4 |
| 3 | 442 | 56.5 | 131 | 12 | US-09-741-873B-2 |
| 4 | 442 | 56.5 | 131 | 12 | US-09-741-873B-2 |
| 5 | 116.5 | 14.9 | 445 | 15 | US-10-369-493-20638 |
| 6 | 105 | 13.4 | 597 | 9 | US-09-793-306-146 |
| 7 | 100 | 12.8 | 943 | 9 | US-09-936-834-131 |
| 8 | 100 | 12.8 | 943 | 10 | US-09-997-182-131 |
| 9 | 100 | 12.8 | 943 | 10 | US-09-997-181-131 |
| 10 | 100 | 12.8 | 943 | 14 | US-10-193-002-199 |
| 11 | 100 | 12.8 | 943 | 14 | US-10-084-843-204 |
| 12 | 100 | 12.8 | 3300 | 12 | US-10-282-122A-64369 |
| 13 | 98.5 | 12.5 | 486 | 15 | US-10-369-493-20619 |
| 14 | 96 | 12.3 | 186 | 12 | US-10-282-122A-49412 |
| 15 | 93 | 11.9 | 892 | 10 | US-09-952-267-5 |

Sequence 114193,
Sequence 62548, A
Sequence 21, Appl
Sequence 193067,
Sequence 162284,
Sequence 23, Appl
Sequence 11, Appl
Sequence 64573, A
Sequence 44999, A
Sequence 62472, A
Sequence 67793, A
Sequence 67750, A
Sequence 10, Appl
Sequence 11, Appl
Sequence 998, App
Sequence 203972,
Sequence 164, App
Sequence 18460, A
Sequence 12420, A
Sequence 64364, A
Sequence 215142,
Sequence 49960, A
Sequence 28, Appl
Sequence 148500,
Sequence 80, Appl
Sequence 105413,
Sequence 152921,
Sequence 120176,
Sequence 148448,
Sequence 55337, A

US-10-437-963-114193
US-10-282-122A-62548
US-09-820-843A-21
US-10-437-963-193067
US-10-437-963-162284
US-10-233-553-23
US-10-233-553-11
US-10-282-122A-64573
US-10-282-122A-44999
US-10-282-122A-62472
US-10-282-122A-67793
US-10-425-114-67750
US-10-185-990-10
US-10-185-990-11
US-10-408-765A-998
US-10-424-599-203972
US-10-479-670-164
US-10-369-493-18460
US-10-369-493-12420
US-10-282-122A-64364
US-10-424-599-215142
US-10-425-114-49960
US-09-810-264-28
US-10-437-963-148500
US-10-479-670-80
US-10-437-963-105413
US-10-437-963-152921
US-10-437-963-120176
US-10-424-599-148448
US-10-425-114-55337

ALIGNMENTS

RESULT 1
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.5%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred No. 2, 2e-44;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALYDQ 60

Db 1 MKLLKVAIAAIFVSGSAVAVFPYGGGNGHGGGNSGENSELNIYQYGGNSALALQ 60
QY 61 LVTRVVTTHMAHAGYGCADVCGQADNSTIELTQGFNNATIDQWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGGADVCGQSDSSIDLQGFNGSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
Db 121 NGRAVDQDTASNSSVNVTVQVGFNNATAHQY 151

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; CURRENT APPLICATION NUMBER: US/09/741,873B
; FILE REFERENCE: 012889-084
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 66.5%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e-44;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAIAAIFVSGSAVAVFPYGGGNGHGGGNSGENSELNIYQYGSANAALYDQ 60
Db 1 MKLLKVAIAAIFVSGSAVAVFPYGGGNGHGGGNSGENSELNIYQYGGNSALALQ 60
QY 61 LVTRVVTTHMAHAGYGCADVCGQADNSTIELTQGFNNATIDQWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGGADVCGQSDSSIDLQGFNGSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
Db 121 NGRAVDQDTASNSSVNVTVQVGFNNATAHQY 151

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020061722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match 56.5%; Score 442; DB 12; Length 131;
Best Local Similarity 64.9%; Pred. No. 1.3e-36;
Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;
QY 21 GVVPOWGGGNGHGGGNSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGCAD 80
Db 1 GVVPOWGGGNGHGGGNSGPDSTLSIYQYGGNSALALQTDARNSDLTITQHGNGGAD 60
QY 81 VQGGADNSTIELTQGFNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVQV 140
Db 61 VQGGSDSSIDLQGFNGSATLDQWNGKNSMTVKQFGGNGAAVDQDTASNSSVNVTVQ 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match 56.5%; Score 442; DB 12; Length 131;

```
Best Local Similarity 64.9%; Pred. No. 1.3e-36;
Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHAGYGNAGD 80
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GWPQYGGGNGHGGGNGGNSGSELNTYQYGGNSALALQTDARNSDLTITQHGNGAGD 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 81 VQGADNSTLTQNGFRNNTATIQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQV 140
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VQGSDDSSIDLQRFGNSATLQWNGKNSMTVKQFGGNGAAVDQTASNSVNTQV 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 141 GFGNNATANQY 151
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GFGNNATAHQY 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233678A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-20638

Query Match 14.9%; Score 116.5; DB 15; Length 445;
Best Local Similarity 27.5%; Pred. No. 0.0032;
Matches 46; Conservative 20; Mismatches 60; Indels 41; Gaps 6;

QY 7 AAPAA-----IVVGSALAGVVPQWGGG-----NHNGG-----GNSGPDSTLSIYQY 50
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 AAPAASNTVYLNQTDGNDQANITQSGNGSVGAFNGSGFLQENGTLSCA-NLLTVKQS 77
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 51 GSANAALYDQLVTRVVTHEMAHAGYNGADVGGADNSTLTQNGFRNNTATIDQWNAKN 110
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 GNSNSVGRD-----IQKQSGAGNSAIFQEGTSGDVELQQTGTSNGAVPSGWNWTN 129
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 111 -----SDITVQYGGNNAALVNQTASDSSVMVRQV 141
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 DPGVFNKITQDSSNSGSKSVIQDGKUNVFSIQKQNTGNSTSVNQIG 176
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
```

```
; TITLE OF INVENTION: of Tuberculosis
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His
US-09-793-306-146

Query Match 13.4%; Score 105; DB 9; Length 597;
Best Local Similarity 29.0%; Pred. No. 0.065;
Matches 40; Conservative 18; Mismatches 58; Indels 22; Gaps 7;

QY 14 VGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 LTGDNLVGIGALNSGIGN-LGFNSG--NNNIGFNSGNNVGVFFNS-----GNNNF 386
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 74 GYNGADVGGADNSTLTQNGFRNNTATIDQ--WNAKNSDITVQYGGNNAALVNQTAS 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 387 GFGNAGDINTGFGNAGD--TNTGFGNAGFNNMGIGNAGNEDMGVNGGSGFNVGVGN--AG 442
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 132 DSSVMVRQVQFGCNNTAN 149
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 443 NOS-----VGFNGAGTLN 455
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-09-996-634-131
; Sequence 131, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

Query Match 12.8%; Score 100; DB 9; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 571 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNI-GFGNVGDY-----NV 610
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 74 GYNGADVGGADNSTLTQNGFRNNTATIDQWNAKNSDITVQYGGN----- 121
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 611 GFGNAGDNQGFANT-----GNNIGFANTGNNNIGIGLSGDNQOQGFNTASGWS 660
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match      12.8%; Score 100; DB 10; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSSGDPSTLSIYQVGSANAALYDQLVTRVVTHEMAHA 73
DB 571 GSGNIGVFNVGSGSLGNVIGSGN-----LGIYNIGFNGVDY-----NV 610

RESULT 8
US-09-997-182-131
; Sequence 131, Application US/09997182
; Publication No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

Query Match      12.8%; Score 100; DB 10; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSSGDPSTLSIYQVGSANAALYDQLVTRVVTHEMAHA 73
DB 571 GSGNIGVFNVGSGSLGNVIGSGN-----LGIYNIGFNGVDY-----NV 610

QY 74 GYGNGADVQGGADNSTIELTQNGFNFNATIDQWNAKSDITVQYGGN-----121
DB 611 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNTASGWS 660

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
DB 661 GTGNSGLFNSGTNNVGFNAGTGNVGIANS GTGN 694

RESULT 9
US-09-997-181-131
; Sequence 131, Application US/09997181
; Publication No. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 61257
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

Query Match      12.8%; Score 100; DB 10; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSSGDPSTLSIYQVGSANAALYDQLVTRVVTHEMAHA 73
DB 571 GSGNIGVFNVGSGSLGNVIGSGN-----LGIYNIGFNGVDY-----NV 610

QY 74 GYGNGADVQGGADNSTIELTQNGFNFNATIDQWNAKSDITVQYGGN-----121
DB 611 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGIGLSGDNQOQGFNTASGWS 660

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
DB 661 GTGNSGLFNSGTNNVGFNAGTGNVGIANS GTGN 694

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/10/193,002
FILING DATE: 10-Jul-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Waki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STREANDEDNESS: <Unknown>
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-10-193-002-199
```

Query Match 12.8%; Score 100; DB 14; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSSGPDSLTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 503

QY 74 GYGNGADVQGGADNSTDITLQNGFRNNATIDQWNAKNSDITVGOYGGN-----121
Db 504 GFGNAGDFNQGFAFNT-----GNNNIGFANTGNNNIGLIGLSDNQOGENIASGWNS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
Db 554 GTGNSGLFNSGTNNVGIFNAGTGNVGIANSCTGN 587

RESULT 11
US-10-084-843-204
; Sequence 204, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonio
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 204:

US-10-084-843-204

Query Match 12.8%; Score 100; DB 14; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSSGPDSLTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 503

QY 74 GYGNGADVQGGADNSTDITLQNGFRNNATIDQWNAKNSDITVGOYGGN-----121
Db 504 GFGNAGDFNQGFAFNT-----GNNNIGFANTGNNNIGLIGLSDNQOGENIASGWNS 553

QY 122 ---NAALVNQTASDSSVM---VRQVGFNNATAN 149
Db 554 GTGNSGLFNSGTNNVGIFNAGTGNVGIANSCTGN 587

RESULT 12
US-10-282-122A-64369
; Sequence 84369, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chlisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64369
; LENGTH: 3300
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64369

Query Match 12.8%; Score 100; DB 12; Length 3300;
Best Local Similarity 26.0%; Pred. No. 1.7;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSSGPDSLTLSIYQGSANAALYDQLVTRVVTHEMAHA 73
Db 590 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 629

QY 74 GYGNGADVQGGADNSTDITLQNGFRNNATIDQWNAKNSDITVGOYGGN-----121
Db 590 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGVDY-----NV 629

```
Db 630 GFGNAGDENQGFANT-----GNNIGFANTGNNIGIGLSDNGQGFNIASGWSN 679
Qy 122 ---NAALVNOTASDSSVM---VRQVGFNNATAN 149
Db 680 GTGNSGLFNSGTTNNVGFNAGTGNVGIANSCTGN 713

RESULT 13
US-10-369-493-20619
; Sequence 20619, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20619
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(486)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20619

Query Match 12.6%; Score 98.5; DB 15; Length 486;
Best Local Similarity 25.7%; Pred. No. 0.23;
Matches 37; Conservative 22; Mismatches 50; Indels 35; Gaps 6;

Qy 11 AIVVSGS-----ALAGVVPQWGGGNGHNGGNSGP-----DSTLSIYQYGSANAAL 57
Db 86 SVWAGTDYRDIIVAGVLPGLGVKKYKAYAGDGGPPFPKDLASDTFFSAPFGADGF- 144
Qy 58 YDQLVTRVVTVMHAGVNGADVGQGDADNSTBELTONGFRNNATIDQWNAKNSDITVG- 116
Db 145 -----VIIHTAAVGRPRGALISQG--NLLI-----AQSSLVDAMRLTEADVNLGM 188
Qy 117 ---QYGGNNAALVNOTASDSSVM 136
Db 189 LPLFHTVGLGLMLTLQQAGGASVI 212

RESULT 14
US-10-282-122A-49412
; Sequence 49412, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49412
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49412

Query Match 12.3%; Score 96; DB 12; Length 186;
Best Local Similarity 29.3%; Pred. No. 0.12;
Matches 49; Conservative 18; Mismatches 62; Indels 38; Gaps 10;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNS-----GPDSTLSIYQYGSANA 56
Db 1 MKL-----SLPATAIAASSLGLVPL-----GAHAADGTISITGTVDSTCSI--NGNANGT 50
Qy 57 LYDQLVTRVVTVMHAGVNGADVG--QGADNST--BELTONG-----FRNNAT 102
Db 51 PADKAIT-----LATVPAGSLASAGAVAGTSNPTDLQLSLTGCTGTATKAIAFPENGPT 104
Qy 103 IDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATAN 149
Db 105 VDOTNGYLSN-TAGTAQNVVEVRLNNAQMPINV---TTGANDITN 147

RESULT 15
US-09-952-267-5
; Sequence 5, Application US/09952267
; Publication No. US2003003272A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-5

Query Match 11.9%; Score 93; DB 10; Length 892;
Best Local Similarity 27.7%; Pred. No. 1.7;
```

| | Matches | 41; Conservative | 14; Mismatches | 49; Indels | 44; Gaps | 9; |
|----|---------|---|----------------|------------|----------|----|
| QY | 28 | GGGNHN-----GGGNSS---GPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGY | 75 | | | |
| Db | 89 | GGGDYNEAKGNYSYTVGGSSNTAKGKSTIGGGDTNDANG-----TVSTIGGGY | 137 | | | |
| QY | 76 | -----GNGADVGGADNSTI--ELTQGFNNATIDQWNAKNSDITVGOYG---GNNAL | 125 | | | |
| Db | 138 | YSRAIGDSSITGGGYNQATGKSTVAGGRNN-----QATGNNSTVAGGSYNOATGNNSTV | 193 | | | |
| QY | 126 | V-----NQTAASDSSVMVRQVGFNNATAN | 149 | | | |
| Db | 194 | AGGSHNQATGEGSF---AAGVENKANAN | 218 | | | |

Search completed: August 2, 2004, 15:36:13
Job time : 37.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782
Sequence: 1 MKLLKVAFAAIVVGSALA.....DSSVMVQVGFNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| ----- | | | | | |

| | | | | | | |
|----|-------|-------|-----|----|---------------------|-------------------|
| 1 | 782 | 100.0 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 2 | 709 | 90.7 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 3 | 692 | 88.5 | 151 | 19 | US-09-543-407-5 | Sequence 57, Appl |
| 4 | 687 | 87.9 | 151 | 6 | US-08-233-642A-57 | Sequence 18, Appl |
| 5 | 675 | 86.3 | 151 | 19 | US-09-543-407-18 | Sequence 28, Appl |
| 6 | 617 | 78.9 | 151 | 19 | US-09-543-407-28 | Sequence 12, Appl |
| 7 | 614 | 78.5 | 151 | 19 | US-09-543-407-12 | Sequence 14, Appl |
| 8 | 612 | 78.3 | 151 | 19 | US-09-543-407-14 | Sequence 24, Appl |
| 9 | 609 | 77.9 | 151 | 19 | US-09-543-407-24 | Sequence 31, Appl |
| 10 | 605 | 77.4 | 131 | 19 | US-09-543-407-31 | Sequence 22, Appl |
| 11 | 601 | 77.4 | 151 | 19 | US-09-543-407-22 | Sequence 30, Appl |
| 12 | 600 | 76.7 | 151 | 19 | US-09-543-407-30 | Sequence 16, Appl |
| 13 | 577 | 73.8 | 151 | 19 | US-09-543-407-16 | Sequence 7, Appl |
| 14 | 523 | 66.9 | 151 | 19 | US-09-543-407-7 | Sequence 4, Appl |
| 15 | 520 | 66.5 | 151 | 13 | US-08-978-878-4 | Sequence 2, Appl |
| 16 | 520 | 66.5 | 151 | 21 | US-09-741-873B-4 | Sequence 2, Appl |
| 17 | 518 | 66.2 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 518 | 66.2 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 507 | 64.8 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 473 | 60.5 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 21 | 466 | 59.6 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 22 | 466 | 59.6 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 23 | 466 | 59.6 | 158 | 30 | US-10-417-886-5834 | Sequence 2, Appl |
| 24 | 442 | 56.5 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 442 | 56.5 | 131 | 21 | US-09-741-873B-2 | Sequence 35, Appl |
| 26 | 338 | 43.2 | 109 | 19 | US-09-543-407-35 | Sequence 37, Appl |
| 27 | 266 | 34.0 | 68 | 19 | US-09-543-407-37 | Sequence 32, Appl |
| 28 | 201.5 | 25.8 | 70 | 19 | US-09-543-407-32 | Sequence 38, Appl |
| 29 | 156 | 19.9 | 48 | 19 | US-09-543-407-39 | Sequence 38, Appl |
| 30 | 116.5 | 14.9 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 31 | 116.5 | 14.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 116.5 | 14.9 | 186 | 30 | US-10-417-886-5833 | Sequence 2, Appl |
| 33 | 116.5 | 14.9 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 34 | 116.5 | 14.9 | 445 | 33 | US-60-360-039-20638 | Sequence 8854, Ap |
| 35 | 114.5 | 14.6 | 145 | 21 | US-09-739-449-8854 | Sequence 8854, Ap |
| 36 | 114.5 | 14.6 | 145 | 23 | US-09-803-110-8854 | Sequence 8854, Ap |
| 37 | 112 | 14.3 | 151 | 19 | US-09-543-407-8 | Sequence 8, Appl |
| 38 | 107 | 13.7 | 151 | 19 | US-09-543-407-6 | Sequence 6, Appl |
| 39 | 105 | 13.4 | 597 | 1 | PCT-US01-05992-146 | Sequence 146, App |
| 40 | 105 | 13.4 | 597 | 22 | US-09-793-306-146 | Sequence 199, App |
| 41 | 100 | 12.8 | 943 | 1 | PCT-US99-03265-199 | Sequence 204, App |
| 42 | 100 | 12.8 | 943 | 1 | PCT-US99-03268-204 | Sequence 199, App |
| 43 | 100 | 12.8 | 943 | 13 | US-08-942-341-199 | Sequence 204, App |
| 44 | 100 | 12.8 | 943 | 13 | US-08-942-578-204 | Sequence 199, App |
| 45 | 100 | 12.8 | 943 | 14 | US-09-024-753-199 | Sequence 199, App |

ALIGNMENTS

RESULT 1
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-26

Query Match 100.0%; Score 782; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.8e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60

QY 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-20
Sequence 20, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 151
TYPE: PRP
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

Query Match 90.7%; Score 709; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 1.3e-67;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAAL--- 57
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 58 -----YDOLVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSD 112
DB 61 SDARKYDOLVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSD 112

QY 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 113 ITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5
Sequence 5, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 151
TYPE: PRP
ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match 88.5%; Score 692; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 8.7e-66;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADVCGADNSTIELTONGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-08-233-642A-57
Sequence 57, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
BASED VACCINES
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7052
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-233-642A-57

Query Match 87.9%; Score 687; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 3e-65;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

```
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNPALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 5
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18

Query Match 86.3%; Score 675; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 5.9e-64; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 45
QY 61 LVTRVVTHEMAHAGYNGADVCGGADNSTIELTQNGFRNNATIDQ 105
Db 46 LVTRVVTHEMAHALQSDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDQ 105
QY 106 WNAKNSDITVGYGNNALVNOTASDSSVMVRQVGFNNATANQY 151
Db 106 WNAKNSDITVGYGNNALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-28

Query Match 78.9%; Score 617; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1e-57; Indels 46; Gaps 2;
Matches 128; Conservative 0; Mismatches 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
QY 58 -----YDQLVTRVVTHEMAHAGYNGADVCGGADNSTIELTQNGF 97
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 97
QY 98 RNNATIDOWNAKNSDITVGYGNNALVNOTASDSSVMVRQVGFNNATANQY 151
Db 98 RNNATIDOWNAKNSDITVGYGNNALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 78.5%; Score 614; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.2e-57; Indels 0; Gaps 0;
Matches 122; Conservative 6; Mismatches 23;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60
QY 61 LVTRVVTHEMAHAGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```

; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match      78.3%; Score 612; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3.6e-57;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYDQ 120

QY 121 NNAALVNOTASDSSVMVROVGFGNNTANQY 151
DB 121 LVTRVVTHEMAHASVMVROVGFGNNTANQY 151

RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      77.9%; Score 609; DB 19; Length 151;
Best Local Similarity 82.1%; Pred. No. 7.5e-57;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVVT 120

```

```

QY 121 NNAALVNOTASDSSVMVROVGFGNNTANQY 151
DB 121 HEWAHANQTSADSSVMVROVGFGNNTANQY 151

RESULT 10
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match      77.4%; Score 605; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.7e-56;
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYNGAD 80
DB 1 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQSDARKSETITQSGYNGAD 60

QY 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROV 140
DB 61 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOTASDSSVMVROV 120

QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

RESULT 11
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match      76.9%; Score 601; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.5e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60

```

```
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 76.7%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 77.7%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7e-56;
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7

Query Match 66.9%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.4e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGCGNGADVGQGSDDSSIDLTQRFNGSATLDQWNGKNSMTVQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 73.8%; Score 577; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2.1e-53;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTTTQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7

Query Match 66.9%; Score 523; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.4e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDPSTLSIYQGSANAALQ 60
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGCGNGADVGQGSDDSSIDLTQRFNGSATLDQWNGKNSMTVQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      66.5%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.9e-47;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGNSSGPDSTLSIYQYGSANAALYDQ 60
Db      1 MKLLKVAIAAIVFSGSAVAGVVPQYGGGNNHGGNNSGPNSEINLYQYGGNSALALQ 60

QY      61 LVTRVVTHEMAHAGYCGADYCGGADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120
Db      61 TDARNSDLTITQHGCGGADYCGGSDSSIDLTQGFNGSATLDQWNGKNSMTVQKQFG 120

QY      121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
Db      121 GNGAAVDQTASNSSVWVTQVGFNNATANQY 151
```

Search completed: August 2, 2004, 15:26:46
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-26
Perfect score: 782
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 520 | 66.5 | 151 | 5 | US-09-741-873C-4 |
| 2 | 442 | 56.5 | 131 | 5 | US-09-741-873C-2 |
| 3 | 93 | 11.9 | 892 | 6 | US-09-952-267B-5 |
| 4 | 93 | 11.9 | 892 | 6 | US-10-872-768-5 |
| 5 | 93 | 11.9 | 892 | 6 | US-10-872-769-5 |
| 6 | 92 | 11.8 | 295 | 6 | US-10-425-115-312468 |
| 7 | 92 | 11.8 | 1459 | 6 | US-10-482-706-129 |
| 8 | 91.5 | 11.7 | 891 | 7 | US-60-566-425-574 |
| 9 | 91.5 | 11.7 | 891 | 7 | US-60-576-812-606 |
| 10 | 91.5 | 11.7 | 894 | 7 | US-60-566-425-570 |
| 11 | 91.5 | 11.7 | 894 | 7 | US-60-566-425-571 |
| 12 | 91.5 | 11.7 | 894 | 7 | US-60-566-425-575 |
| 13 | 91.5 | 11.7 | 894 | 7 | US-60-566-425-578 |
| 14 | 91.5 | 11.7 | 894 | 7 | US-60-576-812-602 |
| 15 | 91.5 | 11.7 | 894 | 7 | US-60-576-812-603 |
| 16 | 91.5 | 11.7 | 894 | 7 | US-60-576-812-607 |
| 17 | 91.5 | 11.7 | 894 | 7 | US-60-576-812-610 |
| 18 | 91.5 | 11.7 | 898 | 6 | US-10-170-205E-12329 |
| 19 | 91.5 | 11.7 | 898 | 6 | US-60-566-425-577 |
| 20 | 91.5 | 11.7 | 898 | 6 | US-60-576-812-609 |
| 21 | 89 | 11.4 | 197 | 6 | US-10-425-115-304391 |
| 22 | 88.5 | 11.3 | 193 | 6 | US-10-425-115-254240 |
| 23 | 88 | 11.3 | 1627 | 6 | US-10-170-205E-16659 |
| 24 | 88 | 11.3 | 1905 | 1 | PCT-US04-09388-9 |
| 25 | 87.5 | 11.2 | 234 | 6 | US-10-767-701-45603 |
| 26 | 87 | 11.1 | 573 | 7 | US-60-565-632-7907 |

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4: Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne

; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4

; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 66.5%; Score 520; DB 5; Length 151;
Best Local Similarity 68.2%; Pred. No. 5.5e-38;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

| | | | | |
|----|-----|---|---------------------------------------|----|
| QY | 1 | MKLLKVAFAAIVVSGSALA | GVVPGVGGGNGHNGGSGDPDSTLSIVQGSANAALYDQ | 60 |
| Db | 1 | MKLLKVAFAAIVVSGSALA | GVVPGVGGGNGHNGGSGNSLNIVQYGGNSALALQ | 60 |
| QY | 61 | LTVRVVTHMAHAGVNGADVCGQAGDNSTIELTQNGFNNATIDQWNAKNSDITVQYGG | 120 | |
| Db | 61 | TDARNSDLTITQHGCGGAGVQGGSDSSIDLITQRFNGSATLDQWNGKNSMTVQYFGG | 120 | |
| QY | 121 | NNAALVNQASDSSVMVQVGFNNATANQY | 151 | |
| Db | 121 | NGAAVDDQTASNSVNTVQVGFNNATAHQY | 151 | |

Query Match 11.9%; Score 93; DB 5; Length 892;
Best Local Similarity 27.7%; Pred. No. 5.7; Mismatches 14; Conservative 41; Indels 44; Gaps 9;
Matches 41; Conservative 14; Mismatches 14; Indels 44; Gaps 9;
QY 28 GGGNHN-----GGGNS--GPDSTLSYQYGSANAALYDQVTRVVTHEVAHAGY 75
DB 89 GGGDYNEAKGNYSTVGGSSNTAKGEKSTIGGDTNDANG-----TYSTIGGY 137
QY 76 -----GNGADVCGGADNSTI--ELTQNGFRNNATIDOWNAKNSDITVGYG--GNNAL 125
DB 138 YSRAIGDSSITGGGYNQATGKSTVAGGRNN-----QATGNNSTVAGGSYNQATGNNSTV 193
QY 126 V----NOTASDSSVMVRQVFGNNATAN 149
DB 194 AGGSHNQATGEGSF---AAGVENKANAN 218
RESULT 4
US-10-872-768-5
; Sequence 5, Application US/10872768
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/10/872,768
; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-768-5
Query Match 11.9%; Score 93; DB 6; Length 892;
Best Local Similarity 27.7%; Pred. No. 5.7; Mismatches 14; Conservative 41; Indels 44; Gaps 9;
Matches 41; Conservative 14; Mismatches 14; Indels 44; Gaps 9;
QY 28 GGGNHN-----GGGNS--GPDSTLSYQYGSANAALYDQVTRVVTHEVAHAGY 75
DB 89 GGGDYNEAKGNYSTVGGSSNTAKGEKSTIGGDTNDANG-----TYSTIGGY 137
QY 76 -----GNGADVCGGADNSTI--ELTQNGFRNNATIDOWNAKNSDITVGYG--GNNAL 125
DB 138 YSRAIGDSSITGGGYNQATGKSTVAGGRNN-----QATGNNSTVAGGSYNQATGNNSTV 193
QY 126 V----NOTASDSSVMVRQVFGNNATAN 149
DB 194 AGGSHNQATGEGSF---AAGVENKANAN 218
RESULT 5
US-10-872-769-5
; Sequence 5, Application US/10872769
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/10/872,769
; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: US/09/336,447

US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2
Query Match 56.5%; Score 442; DB 5; Length 131;
Best Local Similarity 64.9%; Pred. No. 2.8e-31; Mismatches 18; Indels 0; Gaps 0;
Matches 85; Conservative 18; Mismatches 18; Indels 0; Gaps 0;
QY 21 GVVPQWGGGNGHGGSGPDSTLSYQYGSANAALYDQVTRVVTHEVAHAGYNGAD 80
DB 1 GVVPQYGGGNGHGGGNSGPNSELTYYGGNSALALQTDARNSDLTITFHGGNGAD 60
QY 81 VCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGNNALVNCATSDSSVMVRQV 140
DB 61 VCGGSDSSIDITQRFNGNSGNSGPNSELTYYGGNSALALQTDARNSDLTITFHGGNGAD 120
QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131
RESULT 3
US-09-952-267B-5
; Sequence 5, Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267B-5


```
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-872-769-5

Query Match
Best Local Similarity 11.9%; Score 93; DB 6; Length 892;
Matches 41; Conservative 14; Mismatches 49; Indels 44; Gaps 9;

QY 28 GCGNHN-----GCGNSS---GPDSTLSIYQGSANAALYDQLVTRVVTHEMAHAGY 75
DB 89 GGGDYNEAKGNYSTVGGSSNTAKGKSTIGGDTNDANG-----TYSTIGGY 137

QY 76 -----GNGADYQGGADNSTI--ELTQNGFRNATIDQNAKNSDITVQYGV---GNNAL 125
DB 138 YSRAIGSSSTIGGGYNNQATGKSTVAGGRNN-----QATGNNSTVAGGSYNQATGNNSTV 193

QY 126 V-----NOTASDSSVMVRQVGFNNATAN 149
DB 194 AGGSHNQATGGSF---AAGVENKANAN 218

RESULT 6
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pcp
US-10-425-115-312468

Query Match
Best Local Similarity 11.8%; Score 92; DB 6; Length 295;
Matches 37; Conservative 13; Mismatches 40; Indels 56; Gaps 5;

QY 26 WGGGNGHNGGNSGPDST-----LSIYQGSANAALYDQLVTRVVTHEMAHAGYNGA 79
DB 138 YGGGSGSGGGYSSGGYAANGYGVGSGGNYNASGGYS-----GSDGYNGA 186

QY 80 DVGGADNSTIELTQNGFRN-----NATIDQNAKN-----SDITVQYGV 119
DB 187 ASGGYANNLSSGNSNRYNTIGSSDNGTGGNSYNPYAGNYNTGGSSSGTILGEFG 246

QY 120 GNNALYNQATSDSSVMVRQVGFNN 145
DB 247 GG-----GFGN 253

RESULT 7
US-10-482-706-129
; Sequence 129, Application US/10482706
; GENERAL INFORMATION:
; APPLICANT: James, Brian William
; APPLICANT: Marsh, Philip
; APPLICANT: Hampshire, Tobias
; TITLE OF INVENTION: Mycobacterial Antigens Expressed During Latency

; FILE REFERENCE: 1581.1030000
; CURRENT APPLICATION NUMBER: US/10/482.706
; CURRENT FILING DATE: 2004-01-02
; PRIOR APPLICATION NUMBER: PCT/GB02/03052
; PRIOR FILING DATE: 2002-07-04
; PRIOR APPLICATION NUMBER: GB 0116385.6
; PRIOR FILING DATE: 2001-07-04
; PRIOR APPLICATION NUMBER: GB 0123993.8
; PRIOR FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 129
; LENGTH: 1459
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-482-706-129

Query Match
Best Local Similarity 11.8%; Score 92; DB 6; Length 1459;
Matches 38; Conservative 20; Mismatches 76; Indels 18; Gaps 5;

QY 7 AAFAAIVVSGSALAGVVPQGGGHHNGGNSGPDSTLSIYQGSANAALYDQLVTRV 66
DB 722 SGFGNVGSGSGFNFAGNLGNSGFLNVGLTSG-----ILNFGTVSGLYNTSTLGLA 775

QY 67 THEMAHAGYNG-----ADVQGGADNSTIELTQNGFRNATIDQNAKNSDI-----TVGQ 117
DB 776 TSAF-HSGVGNLDSQLAGPMNAAGT--LPNFGANDGTNLGNANLGDYVNGSNGVS 832

QY 118 YGNNALYNQATSDSSVMVRQVGFNNATAN 149
DB 833 YNFGSGNIGNSGFGNIGNSNNFGNFGNVSNN 864

RESULT 8
US-60-566-425-574
; Sequence 574, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566.425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 574
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-574

Query Match
Best Local Similarity 11.7%; Score 91.5; DB 7; Length 891;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHNGGNSGPDSTLSIYQGSANAALYDQLVTRVVTHEMAHAGYCN 77
DB 792 GGGGSDYNEKFNYSGGSGRSGNS-----YSGGASY-----NPGSHGGYGG 835

QY 78 GADVQGGADNSTIELTQNGFRNATIDQNAKNSDITVQYCGNNAALYNQATSDSSVMV 137
DB 836 GSGGG-----SSYQKGQGGYSQS---NYNSPGS---GQ-----NYSGPPSSYSQS 873

QY 138 RQVGFNNNA 146
DB 874 SQGGYGRNA 882

RESULT 9
US-60-576-812-606
; Sequence 606, Application US/60576812
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno
```

us-09-543-407-26.rapn

Tue Aug 3 10:54:43 2004

```
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001527PROV
; CURRENT APPLICATION NUMBER: US/60/576,812
; CURRENT FILING DATE: 2004-06-04
; NUMBER OF SEQ ID NOS: 1501
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 606
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-576-812-606

Query Match      11.7%; Score 91.5; DB 7; Length 891;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 792 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 835

QY 78 GADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 836 GSGGG-----SSYQKGQGYQS-----NYSNPGS-----GQ-----NYSGPPSSYQS 873

QY 138 ROVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 874 SQGGYGRNA 882

RESULT 10
US-60-566-425-570
; Sequence 570, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 570
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-570

Query Match      11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 795 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 838

QY 78 GADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 839 GSGGG-----SSYQKGQGYQS-----NYSNPGS-----GQ-----NYSGPPSSYQS 876

QY 138 ROVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 877 SQGGYGRNA 885

RESULT 11
US-60-566-425-571
; Sequence 571, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 578
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-571

Query Match      11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 795 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 838

QY 78 GADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 839 GSGGG-----SSYQKGQGYQS-----NYSNPGS-----GQ-----NYSGPPSSYQS 876

QY 138 ROVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 877 SQGGYGRNA 885

RESULT 12
US-60-566-425-575
; Sequence 575, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 575
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-575

Query Match      11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

QY 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGN 77
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 795 GGGGSDYNYESKFNSGGRSGGNS-----YSGGASY-----NPGSHGGYGG 838

QY 78 GADVGGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMV 137
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 839 GSGGG-----SSYQKGQGYQS-----NYSNPGS-----GQ-----NYSGPPSSYQS 876

QY 138 ROVGFGNNA 146
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 877 SQGGYGRNA 885

RESULT 13
US-60-566-425-578
; Sequence 578, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 578
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-566-425-578
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 692 | 88.5 | 151 | 2 JC6039 | fimbrin protein ag |
| 2 | 692 | 88.5 | 151 | 2 A10635 | major curlin chain |
| 3 | 523 | 66.9 | 151 | 2 S70788 | curlin protein csg |
| 4 | 501.5 | 64.1 | 152 | 2 D90806 | curlin major subun |
| 5 | 501.5 | 64.1 | 152 | 2 H85665 | hypothetical prote |
| 6 | 114.5 | 14.6 | 145 | 2 AD3143 | conserved hypotet |
| 7 | 114.5 | 14.6 | 145 | 2 H98144 | hypothetical prote |
| 8 | 113 | 14.5 | 2174 | 2 E95965 | hypothetical glyci |
| 9 | 112 | 14.3 | 151 | 2 S70787 | curlin nucleator p |
| 10 | 112 | 14.3 | 151 | 2 C90806 | minor curlin subun |
| 11 | 112 | 14.3 | 151 | 2 G85665 | curlin minor chain |
| 12 | 107.5 | 13.7 | 1322 | 2 S07053 | ice nucleation pro |
| 13 | 107 | 13.7 | 151 | 2 JC6040 | fimbrin protein ag |
| 14 | 107 | 13.7 | 151 | 2 AH0635 | nucleation compone |
| 15 | 106 | 13.6 | 1034 | 2 JC2143 | ice nucleation act |
| 16 | 105.5 | 13.5 | 645 | 2 F70825 | probable PPE prote |
| 17 | 105 | 13.4 | 590 | 2 E70946 | probable PPE prote |
| 18 | 104.5 | 13.4 | 1258 | 2 JQ0188 | ice nucleation pro |
| 19 | 102.5 | 13.1 | 552 | 2 D70604 | probable PPE prote |
| 20 | 102.5 | 13.1 | 1567 | 2 S11672 | ice nucleation pro |
| 21 | 102.5 | 13.1 | 1655 | 2 E97935 | hypothetical prote |
| 22 | 100 | 12.8 | 3300 | 2 D70575 | probable PPE prote |
| 23 | 97.5 | 12.5 | 1028 | 2 A56038 | DNA-binding protei |
| 24 | 97.5 | 12.5 | 1213 | 2 S16356 | ovo protein - frui |
| 25 | 97 | 12.4 | 615 | 2 E70663 | probable PPE prote |
| 26 | 96.5 | 12.3 | 447 | 2 G84687 | probable disease r |
| 27 | 95.5 | 12.2 | 1651 | 2 JC1340 | outer membrane pro |
| 28 | 95 | 12.1 | 652 | 2 E7857 | cell surface anti |
| 29 | 94 | 12.0 | 599 | 2 B42049 | leishmanolysin (EC |

30 94 12.0 599 2 A44951 leishmanolysin (EC
31 93 11.9 434 2 E70768 hypothetical glyci
32 92.5 11.8 575 2 S35327 protein kinase sgg
33 92.5 11.8 639 2 C42049 leishmanolysin (EC
34 92.5 11.8 963 2 B70524 probable PPE prote
35 92.5 11.8 1053 2 B70987 probable PPE prote
36 92 11.8 354 2 B70663 probable PPE prote
37 92 11.8 1436 2 B70520 probable PPE prote
38 92 11.8 1748 2 S42136 cnjB protein - Tet
39 91.5 11.7 582 2 F70675 probable PPE prote
40 91 11.6 678 2 A70762 probable PPE prote
41 90.5 11.6 1200 1 SNFSO ice nucleation pro
42 90.5 11.6 3716 2 E70969 probable PPE prote
43 90 11.5 586 2 T26667 hypothetical prote
44 90 11.5 1390 2 T14004 trfA protein - sli
45 89.5 11.4 1317 2 A54831 nuclear pore compl

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfA precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999

C:Accession: JC6039; PC6015; A44898

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrae.

A:Reference number: JC6039; MUID:96146512; PMID:8550497

A:Accession: JC6039

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714

A:Accession: PC6015

A:Molecule type: protein

A:Residues: 21-52 <CO2>

A:Experimental source: strain 27655-3b

A:Note: the authors translated the codon ACG for residue 44 as Ile

R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A:Title: Purification and characterization of thin, aggregative fimbrae from Salmonell

A:Reference number: A44898; MUID:91310586; PMID:1677357

A:Contents: 27655

A:Accession: A44898

A>Status: preliminary

A:Molecule type: protein

A:Residues: 21-33 <CO3>

A:Note: sequence extracted from NCBI backbone (NCBIP:45936)

C:Genetics:

A:Gene: agfA

C:Function:

A:Description: major component of thin aggregative fimbrae

A:Note: fimbrae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbrin

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match

Best Local Similarity 88.5%; Score 692; DB 2; Length 151;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSGDSTLSIYQGSANALVDQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSGDSTLSIYQGSANALQ 60

QY 61 LVTRVVVTHEMAHAGYNGADVGQGDADNSTIELTQNGFNATIDOWNAKNSDITVGQYGG 120

DB 61 SPARKSETTITCSGNGADVGQGDADNSTIELTQNGFNATIDOWNAKNSDITVGQYGG 120

QY 121 NNAALVNQATSDSSVMVRQVFGNNATANQY 151

Db 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

RESULT 2

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of *csaG*
A:Reference number: S31202; PMID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6, 'V', 8-151 <OLS1>
A:Cross-references: EMBL:U04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42; 44-50 <OLS2>
R:Olson, A.N.; Arngvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133, 'RQDSGWLW' <OLS3>
A:Cross-references: EMBL:U04979; NID:G290424; PIDN:AAA23616.1; PID:G290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: *csaG*
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between *CsaG* and *CsgB* tr
A:Note: curli are thin, coiled fibers expressed on the surface of *Escherichia coli* that
and H-kininogen; in the absence of *CsaG*, *CsgB* can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 88.5%; Score 692; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.6e-51;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVTSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
Db 1 MKLLKVAFAAIVTSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGVNGADVCGQADNSTIELTQNGFRNATIDQWNAKSDITVQYGG 120
Db 61 SPARKSETTITQSGYNGADVCGQADNSTIELTQNGFRNATIDQWNAKSDITVQYGG 120

QY 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

RESULT 3

S70788
curlin protein *csaG* precursor - *Escherichia coli* (strain K-12)
N:Alternate names: *csaG* protein; major curlin protein
C:Species: *Escherichia coli*
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olson, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two *csa* operons is required for production of fibronectin- and C
A:Reference number: S70783; PMID:96414468; PMID:8817489
A:Accession: S70788
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:G1147558; PIDN:CAA62282.1; PID:G1147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; PMID:97426617; PMID:9278503
A:Accession: G64846
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE0000205; GB:U000096; NID:G1787265; PIDN:AACT4126.1; PID:G1787279;
A:Experimental source: strain K-12, substrain MG1655
R:Olson, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

Query Match 64.1%; Score 501.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 3.4e-35;
Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVTSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYD 59
Db 1 MKLLKVAFAAIVTSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60

QY 60 QLTVRVVTHEMAHAGVNGADVCGQADNSTIELTQNGFRNATIDQWNAKSDITVQYGG 119

Db 61 QADARNSDLTITQHGCGNGADVGQGSDDSSIDLTRQFGNSATLQWNGKDSHTVKQFG 120
 QY 120 GNNALVNTQASDSSVMVRQVGFNNATANQY 151
 Db 121 GGNGAAVDQATSNSTVNTQVGFNNATAHQY 152

RESULT 5

H85665
 A:Title: conserved hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H85665
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoumis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H85665
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-152 <STO>
 A:Cross-references: GB:AB005174; NID:gl2514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: csgA

Query Match 64.1%; Score 501.5; DB 2; Length 152;
 Best Local Similarity 67.1%; Pred. No. 3.4e-35;
 Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

QY 1 MLLKVAAPAIIVVSGSALAGVVPQW-GGGNHNGGNSGPDSTLSIYQGSANAALYD 59
 Db 1 MLLKVAAPAIIVVSGSALAGVVPQWGGGGNGGNSGPNSELNIYQGGNSALAL 60
 QY 60 QLVTRVVTHEMAHAGYNGADVGQGDNSTLTQNGFRNNATIDQWNAKNSDITVGYOYG 119
 Db 61 QADARNSDLTITQHGCGNGADVGQGSDDSSIDLTRQFGNSATLQWNGKDSHTVKQFG 120
 QY 120 GNNALVNTQASDSSVMVRQVGFNNATANQY 151
 Db 121 GGNGAAVDQATSNSTVNTQVGFNNATAHQY 152

RESULT 6

AD3143
 A:Title: conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AD3143
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gilet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AD3143
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AAL45562.1; PID:gl7743277; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu4768
 A:Map position: linear chromosome

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAAPAIIVVSGSALAGVVPQWGG-----GNHNGGNSGPDSTLSIYQGSANA 55
 Db 1 MIRKSFIALVALVGLSAAAPAMANDVRIEQVGSNSAGGAQEGYGNRIITYQNGYN- 59
 QY 56 ALYDQLVTRVVTHEMAHAGYNGADVGQGDNSTLTQNGFRNNATIDQWNAKNSDITV 115
 Db 60 -----RIVGHQY---GRHNL SAVQEGHDNYGSTTQNGNRNVAGI----- 96
 QY 116 GOYGGNNAALVNTQASDSSVMVRQVGFNNATANQ 150
 Db 97 GQFGSNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 7

H98144
 A:Title: hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C:Accession: H98144
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourollo, B.; Goldman
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: H98144
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK88682.1; PID:gl5158413; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_228
 A:Map position: linear chromosome

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAAPAIIVVSGSALAGVVPQWGG-----GNHNGGNSGPDSTLSIYQGSANA 55
 Db 1 MIRKSFIALVALVGLSAAAPAMANDVRIEQVGSNSAGGAQEGYGNRIITYQNGYN- 59
 QY 56 ALYDQLVTRVVTHEMAHAGYNGADVGQGDNSTLTQNGFRNNATIDQWNAKNSDITV 115
 Db 60 -----RIVGHQY---GRHNL SAVQEGHDNYGSTTQNGNRNVAGI----- 96

QY 116 GOYGGNNAALVNTQASDSSVMVRQVGFNNATANQ 150

Db 97 GQFGSNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131

RESULT 8

E95965
 A:Title: hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) mag
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95965
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Herna
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95965
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2174 <KUR>
 A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:gl5140875; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001

Query Match 14.6%; Score 114.5; DB 2; Length 145;
 Best Local Similarity 25.2%; Pred. No. 0.0088;
 Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Ielaure
 Hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

• **Chlorophyll** /klo-ro-fil/

821

| |
|----------------------------|
| CCECCNCCACCEWVSTFDSATPECCG |
| : : : : : |
| 794 |

QY
99 -NNATIDQWNAKNSDITVGQYCGGNNALVNQTASDSSVMVRQVGFNNATANQY 151

C;Species: Erwinia ananas

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFQGNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 692 | 88.5 | 151 | 1 | CSGA_SALTY |
| 2 | 523 | 66.9 | 151 | 1 | CSGA_ECOLI |
| 3 | 501.5 | 64.1 | 152 | 1 | CSGA_ECO57 |
| 4 | 112 | 14.3 | 151 | 1 | CSGB_ECOLI |
| 5 | 107.5 | 13.7 | 1322 | 1 | ICEN_PANAN |
| 6 | 107 | 13.7 | 151 | 1 | CSGB_SALTY |
| 7 | 107 | 13.7 | 151 | 1 | CSGB_SALTY |
| 8 | 106 | 13.6 | 1034 | 1 | ICEN_PANAN |
| 9 | 104.5 | 13.4 | 1258 | 1 | ICEN_ERWHE |
| 10 | 102.5 | 13.1 | 1567 | 1 | ICEN_XANCT |
| 11 | 102.5 | 13.1 | 1655 | 1 | ONPB_RICCN |
| 12 | 97.5 | 12.5 | 1028 | 1 | OVG_DROME |
| 13 | 96.5 | 12.3 | 1196 | 1 | ICEV_PSEEX |
| 14 | 95.5 | 12.2 | 1656 | 1 | ONPB_RICUA |
| 15 | 94 | 12.0 | 599 | 1 | GP63_LEICH |
| 16 | 93 | 11.9 | 491 | 1 | YK98_MYCTU |
| 17 | 91.5 | 11.7 | 894 | 1 | IFP3_HUMAN |
| 18 | 91 | 11.6 | 678 | 1 | YF48_MYCTU |
| 19 | 90.5 | 11.6 | 1200 | 1 | ICEN_PSEEX |
| 20 | 89.5 | 11.4 | 1317 | 1 | N145_YEAST |
| 21 | 88.5 | 11.3 | 646 | 1 | GP63_LEIME |
| 22 | 88.5 | 11.3 | 1148 | 1 | ICEK_PSEEX |
| 23 | 88.5 | 11.3 | 1210 | 1 | ICEN_PSEFL |
| 24 | 87.5 | 11.2 | 602 | 1 | GP63_LEIMA |
| 25 | 86.5 | 11.1 | 590 | 1 | GP63_LEIDO |
| 26 | 86 | 11.0 | 959 | 1 | N100_YEAST |
| 27 | 85 | 10.9 | 681 | 1 | YDHE_SCHPO |
| 28 | 84.5 | 10.8 | 548 | 1 | CEAK_ECOLI |
| 29 | 84.5 | 10.8 | 641 | 1 | IMD_ARTGO |
| 30 | 84.5 | 10.8 | 1300 | 1 | 120K_RICRI |
| 31 | 84.5 | 10.8 | 1654 | 1 | ONPB_RICRI |
| 32 | 82.5 | 10.5 | 487 | 1 | Y442_MYCTU |
| 33 | 82 | 10.5 | 1327 | 1 | TNKL_HUMAN |

34 81.5 10.4 347 1 MSA2_PLAF2
35 81.5 10.4 1778 1 N189_SCHPO
36 81 10.4 365 1 ROAL_DROME
37 81 10.4 443 1 Y878_MYCTU
38 81 10.4 877 1 SEU_ARATH
39 80.5 10.3 1063 1 SPT5_YEAST
40 80 10.2 1293 1 MLE_DROME
41 79.5 10.2 1067 1 SGG_DROME
42 79 10.1 165 1 GRP1_ORVSA
43 79 10.1 346 1 RO21_XENLA
44 79 10.1 1302 1 FRPA_NEIMB
45 79 10.1 1645 1 ONPB_RICTY

Q03646 plasmodium
Q0utk4 schizosacch
P07909 drosophila
Q0540 mycobacteri
Q8w234 arabidopsis
P27692 saccharomyc
P24785 drosophila
P18431 drosophila
P25074 oryza sativ
P51989 xenopus lae
Q9X0k9 neisseria m
P96989 r outer mem

ALIGNMENTS

RESULT 1
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEFI7).
DE CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
GN Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
[1] _TaxID=602, 601, 592;
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=SR-11;
RC MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feitwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [4]
SEQUENCE FROM N.A.
RP SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RX Deng W., Liu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

"Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.";
 J. Bacteriol. 185:2330-2337(2003).
 [5] SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6] SEQUENCE OF 21-151 FROM N.A.
 RP SPECIES=S.enteritidis; STRAIN=27655-3B;
 RC SPECIES=S.enteritidis; PubMed=8104955;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,
 Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7] SEQUENCE OF 21-33.
 RP SPECIES=S.enteritidis; STRAIN=27655-3B;
 RC MEDLINE=91310586; PubMed=1677357;
 RX Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RA "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AJ002301; CA05317.1; -
 DR EMBL; AE008749; AAL20074.1; -
 DR EMBL; AL627269; CAD08268.1; -
 DR EMBL; AE016840; AAO89399.1; -
 DR EMBL; U43280; AAC43599.1; -
 DR FIR; JC6039; JC6039.
 DR StyGene; SG10608; csGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
 FT CONFLICT 134 151 SVMYRVQFGNNATANY -> DSYTOVAS (IN
 REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 Query Match 88.5%; Score 692; DB 1; Length 151;
 Best Local Similarity 90.7%; Pred. No. 3.2e-51;
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
 Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQVGSNAALYDQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNNSGPDSTLSIYQVGSNAALQ 60
 Qy 61 LVTRVVTHEAHAGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 Db 61 SDAREKSTTTQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
 Qy 121 NNAALVNQTSADSSVMYRVQFGNNATANY 151
 Db 121 NNAALVNQTSADSSVMYRVQFGNNATANY 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID _CSGA_ECOLI
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 CX NCBI_TaxID=562;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csGA, the subunit gene of fibronectin-binding curli in
Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2] SEQUENCE FROM N.A.
 RP STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csG operons is required for production of
 fibronectin- and congo red-binding curli polymers in *Escherichia coli*
 K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3] SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4] SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:1137-155(1996).
 RN [5] SEQUENCE OF 21-40.
 RP STRAIN=K12 / YMWEL;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 fibronectin binding in *Escherichia coli* HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6] SEQUENCE OF 21-31.
 RP MEDLINE=91310586; PubMed=1677357;
 RX Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.

RT K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of *Escherichia coli* K-12.";
 RT Science 277:1453-1474 (1997).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RC MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=95157246; PubMed=7854117;
 RA Arngvist A., Olsen A., Normark S.;
 RA "Sigma S-dependent growth-phase induction of the *csgB* promoter in
 RT *Escherichia coli* can be achieved in vivo by sigma 70 in the absence
 RT of the nucleoid-associated protein H-NS.";
 RL Mol. Microbiol. 13:1021-1032(1994).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X90754; CAA62281.1; -.

DR EMBL; AE002025; AAC74125.1; -.
 DR EMBL; D90741; BAA35831.1; -.
 DR EMBL; AE005315; BAG55787.1; -.
 DR EMBL; AP002554; BAB34842.1; -.
 DR PIR; G90806; C90806.
 DR PIR; G85665; G85665.
 DR PIR; S70787; S70787.
 DR EcoGene; EG12621; csgB.
 KW Fibria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
 Query Match 14.3%; Score 112; DB 1; Length 151;
 Best Local Similarity 29.8%; Pred. No. 0.0083;
 Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;
 Qy 51 GSNAALYDQLVT--RVVTHEAHAGYNGADVGGQADNSTIELTQNGFR----- 98
 Db 17 GTAAAGYDLANSEYNFAVNELSKSFNOAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76
 Qy 99 -NNATIDQWNAKNSDITVGYGNGNAALVNQATSDSSVVMVGVFGNNATANQY 151
 Db 77 SNRAKIDQTDYNL-AYIDQASANDASISQAYGNTAMIIQGGSKNKANITQY 129
 RESULT 5
 ICEA PANAN STANDARD; PRT; 1322 AA.
 AC P20469;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inaa.
 GN INAA.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9092494; PubMed=2599095;
 RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
 RA "An ice nucleation active gene of *Erwinia ananas*. Sequence similarity
 RT to those of *Pseudomonas* species and regions required for ice
 RT nucleation activity.";
 RL FEBS Lett. 258:297-300(1989).
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
 CC crystallization in supercooled water.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
 CC family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X17316; CAA35194.1; -.
 DR PIR; S07053; S07053.
 DR EMBL; P06620; IINA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleation; 69.
 DR PRINTS; PS00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 49.
 KW Ice nucleation; Repeat; Outer membrane.
 FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.

SQ SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;
Query Match 13.7%; Score 107.5; DB 1; Length 1322;
Best Local Similarity 29.9%; Pred. No. 0.22;
Matches 38; Conservative 19; Mismatches 21; Indels 49; Gaps 8;
QY 34 GGNSSGPDSTLSYQYGSANAALYDQVTRVTHMAHAGYNGADVGQADNSTIELT 93
DB 933 GSTTAGPDSL-IGAGYSTQTATGNSILT-AGYGS-----T 967
QY 94 QNGFRNATIDQWAKNSDITVGOY-----GNNAALV-----NOTASDSSVMVRQVGF 143
DB 968 QTG-----QENSDLTTG-YGSTAGYESSLIAGVSTQTATGNSILT-AGY 1013
QY 144 NNATANQ 150
DB 1014 SSQTARE 1020
RESULT 6
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Baham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dord L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holloway S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.,
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RN J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DB EMBL; AL627269; CAD08267.1; -;
DB EMBL; AF016840; A069400.1; -;

KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
Query Match 13.7%; Score 107; DB 1; Length 151;
Best Local Similarity 30.4%; Pred. No. 0.22; Indels 16; Gaps 5;
Matches 35; Conservative 17; Mismatches 47;
QY 51 GSANAALYD--QLVTRVTHMAHAGYNGADVGQ--GADNST-----IELTQNGF 97
DB 17 GIATATNYDLARSEYNFAVNSLSKSFNQAAIIQGVGTDSARVRQEGSKLLSVISQEG 76
QY 98 RNATIDQWAKNSDIT-VGOYGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 NNRKVDQ--AGNYNFAYIEQTGNANDASISOSAYGNSAAIIQKSGNKANITQY 129
RESULT 7
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RN J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=enteritidis; STRAIN=27655-3B;
RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";
RN J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DB EMBL; AL627269; CAD08267.1; -;
DB EMBL; AF016840; A069400.1; -;

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; AJ002301; CAA05316.1; -;
 DR EMBL; AB008749; AAL20073.1; -;
 DR EMBL; U43280; AAC43598.1; -;
 DR PIR; JC6040; JC6040.
 DR StyGene; SG10609; csGB.
 KW Fimbrtia; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16182 MW; CFCF5430E6DD361D CRC64;
 Query Match 13.7%; Score 107; DB 1; Length 151;
 Best Local Similarity 30.4%; Pred. No. 0.022;
 Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;
 QY 51 GSANAALYD--QLVTRVVTHEMAHAGYNGADVGO--GADNST-----IELTQNGP 97
 DB 17 GIATATNYDLARSEYFNFAVNEKSGSFNQAAIIGVGTDSARVRQEGSKLLSVISQEGG 76
 QY 98 RNNATIDOWNAKNSDIT-VGOYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 77 NNRKVDQ--AGNYNFAVIEQTGNANDASISQSAVNSAAIIQKSGNKANITQY 129
 RESULT 8
 ICEN_PANAN STANDARD; PRT; 1034 AA.
 ID ICEN_PANAN
 AC Q47879;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein inau.
 GN INAU.
 OS Pantoea ananas (Erwinia uredovora).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=553;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KUN-3;
 RX MEDLINE=94264407; PubMed=7764866;
 RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia uredovora";
 RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate crystallization in supercooled water.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; D14992; BAA03636.1; -;
 DR PIR; J22143; J22143.
 DR HSP; P06620; IINA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleatn; 51.
 DR PRINTS; PR00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 34.
 KW Ice nucleation; Repeat; Outer membrane.

FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
 SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;
 Query Match 13.6%; Score 106; DB 1; Length 1034;
 Best Local Similarity 29.9%; Pred. No. 0.22;
 Matches 40; Conservative 19; Mismatches 37; Indels 38; Gaps 9;
 QY 27 GCGGNHNGGNSGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYNGADVGOQAD 86
 DB 531 GYGSTAGANSS-----LIAGYGSTQTASVNSVLT-----AGYGSTQTAREGSD 575
 QY 87 NSTIELTQNGFRNNATIDOWNAKNSDITVGOYQ-----GNNAALV-----NOTASDSSVM 136
 DB 576 -----LTA-GYGSTQTAQE-----NSDLTTG-YGS*STAGYDSSLIAGYGSTQTAGYHSIL 624
 QY 137 VRQVGFNNATANQ 150
 DB 625 T--AGYGSTQTAQE 636
 RESULT 9
 ICEN_ERWHE STANDARD; PRT; 1258 AA.
 ID ICEN_ERWHE
 AC P16239;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN ICEE.
 OS Erwinia herbicola.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M1;
 RX MEDLINE=90152370; PubMed=2515997;
 RA Warren G.J., Corotto L.V.;
 RT "The consensus sequence of ice nucleation proteins from Erwinia herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
 RL Gene 85:239-242(1989).
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate crystallization in supercooled water.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; M26382; AAA24823.1; -;
 DR PIR; JQ0188; JQ0188.
 DR HSP; P06620; IINA.
 DR InterPro; IPR000258; Ice_nucleatn.
 DR Pfam; PF00818; Ice_nucleatn; 65.
 DR PRINTS; PR00327; ICENUCLEATN.
 DR PROSITE; PS00314; ICE_NUCLEATION; 45.
 KW Ice nucleation; Repeat; Outer membrane.
 FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
 SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;
 Query Match 13.4%; Score 104.5; DB 1; Length 1258;
 Best Local Similarity 29.5%; Pred. No. 0.36;


```
Matches 44; Conservative 23; Mismatches 37; Indels 45; Gaps 10;
QY 34 GGNSSGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYN-----GA 79
Db 773 GSTTAGADSSL-IAGYGSTQTAGYHSILT-----AGYGSTQTAQERSDLTTGYGS 822
QY 80 DVGQADNSTIE---LTQN-GFRNNATI-----DQWNAKNSDITVQYQ-----GNNAAIV 126
Db 823 TSTAGADSSLIAGYGSTQTAGYNSILT-IAGYGSTQTAQNSDLTTG-YGSTTAGYDSSLI 881
QY 127 -----NOTASDSSVMVRQVGFNNATANQ 150
Db 882 AGYGSTQTAGYNSILT--AGYGSTQTAQE 908

RESULT 10
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X565;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52970; CAA37140.1; -.
CC HSPB; P06620; 1INA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 81.
CC PRINTS; PR00327; ICNUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 57.
CC Ice_nucleation; Repeat; Outer membrane.
CC SEQUENCE 1567 AA; 152548 MW; C8B451D9595ECAD63 CRC64;

Query Match
Best Local Similarity 30.2%; Pred. No. 0.68;
Matches 45; Conservative 23; Mismatches 36; Indels 45; Gaps 11;
QY 34 GGNSSGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYN-----GADV----- 81
Db 1055 GSTGTAGADSTL-IAGYGSTQTAGSDSLT-----AGYGSTQTAQSGSDITAGYGS 1104
QY 82 --GQADNSTIE---LTQN-GFRNNATI-----DQWNAKNSDITVQYQ-----GNNAAIV 126
```

```
Db 1105 TOTAGADSSLIAGYGSTQTAGYDSNLTAGYGSTQTAQREDSLITAG-YGSTTAGDSSLI 1163
QY 127 -----NOTASDSSVMVRQVGFNNATANQ 150
Db 1164 AGYGSTQTAGYNSILT--TG YGSTQTAQE 1190

RESULT 11
ID OMPB_RICCN STANDARD; PRT; 1655 AA.
AC Q9KK3; O9KK98; O9KC45;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB OR RC1085.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Malish 7;
RX MEDLINE=2142074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [2]
RP SEQUENCE OF 33-1649 FROM N.A.
RX STRAIN=Indian tick typhus, and Malish 7;
RX MEDLINE=20393643; PubMed=10939649;
RA Roux V.; Raoult D.;
RT "Phylogenetic analysis of members of the genus Rickettsia using the
RT gene coding the outer-membrane protein rOmpB (ompB).";
RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN [3]
RP SEQUENCE OF 353-1655 FROM N.A.
RX STRAIN=Malish 7;
RA Stenos J., Walker D.;
RT "The rickettsial outer membrane protein A and B genes of Rickettsia
RT australis, the most divergent rickettsia of the spotted fever group.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DDAT databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; A5008659; AAL03623.1; -.
CC EMBL; AF123721; AAF34124.1; -.
CC EMBL; AF123726; AAF34129.1; -.
CC EMBL; AF149110; AAD39533.1; -.
CC PIR; E97835; E97835.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
```



```

ID AC ICEV PSEX STANDARD; PRT; 1196 AA.
AC O33479;
DE 15-DEC-1998 (Rel. 37, Created)
DE 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAV.
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INA5;
RX MEDLINE=97462815; PubMed=9323042;
RA Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
RA Jann A.
RT "Molecular organisation of the ice nucleation protein Inav from
RT Pseudomonas syringae."
RL FEBS Lett. 414:590-594 (1997).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
CC FOR ICE NUCLEATION ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001086; CAA04521.1; -
CC HSP; P06620; LINA.
CC InterPro; IPR000258; Ice nucleatn.
CC Pfam; PF00818; Ice nucleation; 61.
CC PRINTS; PR00327; ICENUCLENTN.
CC PROSITE; PS00314; ICE_NUCLEATION; 42.
CC Ice nucleation; Repeat; Outer membrane.
CC DOMAIN 172 1147 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1196 AA; 117991 MW; C9E9974CB1731E68 CRC64;
Query Match 12.3%; Score 96.5; DB 1; Length 1196;
Best Local Similarity 26.7%; Pred. No. 1.6;
Matches 43; Conservative 24; Mismatches 39; Indels 55; Gaps 11;
QY 16 GSALAG-----VVPQWGG---GNHN-----GGNSSGPDSTLSIYQYGSANAALYDQL 61
Db 175 GSTLSGDNNSRLIAGYGSNTAGNSDLIAGYGSTGTAGSDSL-VAGYGSTGTAGGDSA 233
QY 62 VTRVVTHEMAHAGYGN-----GADVGGAGDNSTLTQTNGFRNNATIDQWN 107
Db 234 LT-----AGVGSQTAREGNSLTAGYGSTGTAGSDSLIA-----GYGSTQT----S 276
QY 108 AKNSDITVGYGGNNAALVNQTSADSSVMVRQVFGNNATA 148
Db 277 GEDSSLTAG-YGS-----TQTAGEGSLNLT--AGYGSTGTA 308
RESULT 14
OMPB_RICUA STANDARD; PRT; 1656 AA.
ID OMPB_RICUA
AC O06653;
DE 30-MAY-2000 (Rel. 39, Created)
DE 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RX Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
CC Pfam; PF03797; Autotransporter; 1.
CC TIGRFAMs; TIGR01414; autotrans_bar1; 2.
CC Antigen; S-layer; Cell wall.
CC CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
CC CHAIN 1339 1656 32 kDa BETA PEPTIDE.
CC DOMAIN 528 533 POLY-GLY.
CC SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;
Query Match 12.2%; Score 95.5; DB 1; Length 1656;
Best Local Similarity 25.0%; Pred. No. 2.8;
Matches 36; Conservative 21; Mismatches 52; Indels 35; Gaps 7;
QY 28 GGGNNGGNGSGSGPDSTLS-----IVQYGSANAALYDQLVTR-----VVTHEMAHAGYGN 77
Db 1215 GGARNF--GTGGPNFTVIGSNRFVNYGLIRAAQDYVITRTNNAENIVTNDITNSPFG 1272
QY 78 GADVGGAGDNSTLTQTNGFRNNATIDQWNAKNSDITVGYGGNNAALVNQTSADSSVMV 137
Db 1273 APGVQGVN-TTFVNATNTAAYNNLL-----AKNS-----ADSANFVGIVTDTSAAI 1319
QY 138 R-----QVGFNNATANQY 151
Db 1320 TNAQLDVAKD-QAQLGNRLGALRY 1343
RESULT 15
GP63_LEICH STANDARD; PRT; 599 AA.
ID GP63_LEICH
AC P15706;
DE 01-APR-1990 (Rel. 14, Created)
DE 01-APR-1990 (Rel. 14, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.

```

OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
CX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205976; PubMed=2320059;
RA Miller R.A., Reed S.G., Parsons M.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an
RT Arg-Gly-Asp sequence."
RL Mol. Biochem. Parasitol. 39:267-274 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
RA Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially
RT expressed during development of Leishmania donovani chagasi
RT promastigotes to an infectious form."
RL J. Biol. Chem. 267:1888-1895 (1992).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1' and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at Ala-Tyr-Leu-Lys-Lys-
CC -!- COPACITOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80672; AA29238.1; -;
DR EMBL; M28527; AA29235.1; -;
DR PIR; A44951; A44951.
DR HSP; P08148; LML.
DR MEROPS; M08.001; -;
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8_1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 122 139 BY SIMILARITY.
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 393 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 227 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By
FT similarity).
SQ SEQUENCE 599 AA; 63848 NW; 746730AE8E2A2E7C CRC64;
Query Match 12.0%; Score 94; DB 1; Length 599;
Best Local Similarity 54.8%; Pred. No. 1.2;

Matches 23; Conservative 5; Mismatches 8; Indels 6; Gaps 2;
QY 52 SAN-AALYDQLVTRVVTTHMAHA-----GYNGADVGQADN 87
Db 244 ANIASRYDQLVTRVVTTHMAHALGFSVGFEGARILEISIN 285
Search completed: August 2, 2004, 14:49:32
Job time : 6.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)

1604.150 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DGSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 683 | 87.3 | 152 | Q33802 | Q33802 salmonella |
| 2 | 580.5 | 74.2 | 150 | Q7X243 | Q7x243 citrobacter |
| 3 | 532 | 68.0 | 149 | Q7X240 | Q7x240 citrobacter |
| 4 | 508.5 | 65.0 | 152 | Q8CW63 | Q8cw63 escherichia |
| 5 | 442.5 | 56.5 | 150 | Q7X237 | Q7x237 enterobacte |
| 6 | 304 | 38.9 | 76 | Q54069 | Q54069 salmonella |
| 7 | 122 | 15.6 | 29 | Q9S3U5 | Q9s3j5 escherichia |
| 8 | 121 | 15.5 | 502 | Q8EIH4 | Q8eih4 shewanella |
| 9 | 114.5 | 14.6 | 145 | Q8UGN9 | Q8u6n9 agrobacteri |
| 10 | 113 | 14.5 | 2174 | Q92UJ8 | Q92uu8 rhizobium m |
| 11 | 112 | 14.3 | 151 | Q7UCZ1 | Q7ucz1 shigella fl |
| 12 | 112 | 14.3 | 160 | Q8CW64 | Q8cw64 escherichia |
| 13 | 112 | 14.3 | 160 | Q8SRU7 | Q8sr7 shigella fl |
| 14 | 110 | 14.1 | 151 | Q7X244 | Q7x244 citrobacter |
| 15 | 108.5 | 13.9 | 1333 | Q8PD38 | Q8pd38 xanthomonas |
| 16 | 107 | 13.7 | 1422 | Q8EFU3 | Q8efu3 shewanella |

| | | | | | | |
|----|-------|------|------|----|--------|--------------------|
| 17 | 105.5 | 13.5 | 645 | 16 | Q7U1C5 | Q7u1c5 mycobacteri |
| 18 | 105.5 | 13.5 | 646 | 16 | O53818 | O53818 mycobacteri |
| 19 | 105 | 13.4 | 589 | 16 | Q7TX53 | Q7tx53 mycobacteri |
| 20 | 105 | 13.4 | 590 | 16 | O53309 | O53309 mycobacteri |
| 21 | 105 | 13.4 | 603 | 16 | Q8VJ66 | Q8vj66 mycobacteri |
| 22 | 104.5 | 13.4 | 151 | 2 | Q7X238 | Q7x238 enterobacte |
| 23 | 104 | 13.3 | 490 | 16 | Q8EYJ9 | Q8eyj9 leptospira |
| 24 | 103.5 | 13.2 | 91 | 2 | Q9S3J8 | Q9s3j8 escherichia |
| 25 | 103.5 | 13.2 | 1306 | 2 | Q93N36 | Q93n36 pantoea ana |
| 26 | 102.5 | 13.1 | 552 | 16 | P96840 | P96840 mycobacteri |
| 27 | 102.5 | 13.1 | 552 | 16 | Q7TW76 | Q7tw76 mycobacteri |
| 28 | 102.5 | 13.1 | 623 | 16 | Q8VJF0 | Q8vjf0 mycobacteri |
| 29 | 102.5 | 13.1 | 1616 | 2 | Q9KKA1 | Q9kkal rickettsia |
| 30 | 102 | 13.0 | 130 | 16 | Q8GJ14 | Q8gji4 bradyrhizob |
| 31 | 102 | 13.0 | 139 | 16 | Q8EIH3 | Q8eih3 shewanella |
| 32 | 101.5 | 13.0 | 152 | 2 | Q7X241 | Q7x241 citrobacter |
| 33 | 101.5 | 13.0 | 1616 | 2 | Q9KKA5 | Q9kka5 rickettsia |
| 34 | 101 | 12.9 | 157 | 16 | Q88H30 | Q88hg0 pseudomonas |
| 35 | 100.5 | 12.9 | 1616 | 2 | Q9KKA9 | Q9kka9 rickettsia |
| 36 | 100 | 12.8 | 191 | 10 | Q7XDR3 | Q7xdr3 oryza sativ |
| 37 | 100 | 12.8 | 3275 | 16 | Q8VKM3 | Q8vkm3 mycobacteri |
| 38 | 100 | 12.8 | 3300 | 16 | O06304 | O06304 mycobacteri |
| 39 | 100 | 12.8 | 3507 | 16 | Q7U270 | Q7u270 mycobacteri |
| 40 | 99.5 | 12.7 | 1613 | 2 | Q84OU5 | Q84ou5 rickettsia |
| 41 | 98.5 | 12.6 | 154 | 16 | Q89J15 | Q89j15 bradyrhizob |
| 42 | 98.5 | 12.6 | 453 | 5 | Q9N6M6 | Q9nem6 drosophila |
| 43 | 98.5 | 12.6 | 738 | 5 | O02402 | O02402 pinctada fu |
| 44 | 98.5 | 12.6 | 1286 | 2 | Q84IY5 | Q84iy5 campylobact |
| 45 | 98.5 | 12.6 | 3659 | 16 | Q98LN6 | Q98ln6 rhizobium 1 |

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.

AC O33802;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Agfa protein (Fragment).
GN AGFA.

OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;

RN [1]_SEQUENCE FROM N.A.

RP MEDLINE=98053981; PubMed=9393832;

RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RL cells";

RL Infect Immun. 65:5320-5325(1997).

DR EMBL; AJ000514; CAA04151.1; -.

FT NON_TER 152 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.3%; Score 683; DB 2; Length 152;

Best Local Similarity 89.4%; Pred. No. 2.8e-48;

Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHNGSGSPDSTLSTIYQYGSANAALYDQ 60
Db 1 MKLLKVAFAAIVVSGSAVAGVVPQWGGGNNHNGSGSPDSTLSTIYQYGSANAALQ 60

QY 61 LVTRVVTTHMAHAGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

QY 121 NNAALVNQATSDSSVMVRQVFGNNATANQY 151


```
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CA056678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8B2D872DF15F3 CRC64;

Query Match      56.6%; Score 442.5; DB 2; Length 150;
Best Local Similarity 60.3%; Pred. No. 1.2e-28;
Matches 91; Conservative 25; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPOMGGNGGNGGNSGPDSTLSIYQYGSANAALYDQ 60
DQ 1 MKFKVAAALAAIVVSGSAGAMGNIQ-GGWGHGHGGYGGPNTLNIYQGGNSALALQ 59

QY 61 LVTRVVTWTHMAHAGYNGGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
DQ 60 TDARNSVLNT-SQTGGGNGADVGQSDSSINLTQNGFGNSATLDQWNSKDSVMNVQYGG 119

QY 121 NNAALVNTASDSSVMVQVGFNNATQY 151
DQ 120 LNALVDQTASNSTVNTQTIGFNGHATAHQY 150

RESULT 6
Q54069      PRELIMINARY;      PRT;      76 AA.
ID Q54069
AC Q54069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -.
DR NON TER 1
DR NON TER 76
FT NONTER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match      38.9%; Score 304; DB 2; Length 76;
Best Local Similarity 80.3%; Pred. No. 1e-17;
Matches 61; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 30 GNHGGGNSGPDSTLSIYQYGSANAALYDQVTRVVTWTHMAHAGYNGGADVGQADNST 89
DQ 1 GNHGGGNSGPDSTLSIYQYGSANAALYDQVTRVVTWTHMAHAGYNGGADVGQADNST 60

QY 90 IELTQNGFRNNATIDQ 105
DQ 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q9S3J5      PRELIMINARY;      PRT;      29 AA.
ID Q9S3J5
AC Q9S3J5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
```

```
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curlation of Escherichia coli O78:k80 isolates associated with
RT IS1 inserti on in cs9B and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -.
DR NON TER 29
FT NONTER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match      15.6%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0022;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPOMGGG 29
DQ 1 MKLLKVAALAAIVFSGSALAGVVPQYGG 29

RESULT 8
Q8EIH4      PRELIMINARY;      PRT;      502 AA.
ID Q8EIH4
AC Q8EIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SO0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umavam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Kouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL: AB015532; AAN53941.1; -.
DR TIGR: SO0865; -.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match      15.5%; Score 121; DB 16; Length 502;
Best Local Similarity 26.1%; Pred. No. 0.086;
Matches 40; Conservative 22; Mismatches 57; Indels 34; Gaps 6;

QY 29 GGNHNG-----GGN-----SSGPDSTLSIYQYGSANAALYDQVTR 64
DQ 231 GDNHTGFVYALAGSNDISMEQEGSNNTAYLSMTTGGDNTVDITQDGSN-TVGDSLAD 289

QY 65 VVTHE---MAHAGYNGGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
DQ 290 IOGDDNDITIKQKGDNGAEQVWGDSDNDVLKQKGDANFATFGAYGTDN-DFDLSXGD 348

QY 121 NNAALVNTASDSSVMVQVGFN-----NATAN 149
DQ 349 NNELVAFATGEDNSIEISQEGDANFAYVDATGN 381

RESULT 9
```

```

OSU6N9
ID Q8U6N9 PRELIMINARY; PRT; 145 AA.
AC Q8U6N9;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein Atu4768.
GN ATU4768 OR AGR L.228.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Strubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Lingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nestor E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmliel C., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009405; AAL45562.1; -.
DR EMBL; AE008209; AAK38682.1; -.
DR PIR; AD3143; AD3143.
DR PIR; H98144; H98144.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 145 AA; 14994 MW; DEDC870E1713D51A CRC64;

Query Match 14.6%; Score 114.5; DB 16; Length 145;
Best Local Similarity 25.2%; Pred. No. 0.064;
Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAFAIVVSSALAGVVPQWG-----CGNNGGNSGSPDSTLSIYQGSANA 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MIRKSFIAISALVGLVLSAAAPAVANDVRIDYQWMSNAGAGQGYGNRTYQNGYN- 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 ALYDQLVPRVVTTHMAHAGYNGADVGGADNNTIELTQNGFRNNAIDQWNAKNSDITV 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 -----RIVGHQY---GRNLSAVGQEGHDNYGSTTQNGNRNVAGI----- 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 GOYGGNNAA--LVNQTSADSSVMVRQVG 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 97 GQFGSNHTTILTDQNGNIAAGVQVGRGCSANVSQ 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
Q92U08 PRELIMINARY; PRT; 2174 AA.
AC Q92U08;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical glycine-rich protein Smb21548.
GN RB0989 OR Smb21548.

OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vornheeler F.J., Hernandez-Lucas I., Becker A., Cowie J.,
RA Golding B., Puchler A.;
RA "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -.
DR PIR; E95965; E95965.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00583; PFKB_KINASES_1; 2.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 2174 AA; 203314 MW; 008BB68297B44182 CRC64;

Query Match 14.5%; Score 113; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 2.3;
Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

QY 11 AIWVGSGALAGVVPQ--WGGGNGHNGGNSGSPDSTLSIYQGSANAALYDQLVTRVVT 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 693 AIATAGAGAGVILAQSIGGGN--CGNATCGDAGFGSGFTGGGGG----- 737
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 EMHAGYNGADVGG-----QGADNSTI--ELTQNGFRNNAIDQWNAK--NSDITV 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 738 ----GGYANTANVGFKGLTLTQGSAAAGIVAQSVGGGGTGTASSIGFTASVAV 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 GOYGGNNAA--LVNQTSADSSVMVRQVG 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 794 GGTGGNGGAGGEVSLSLTDSAIRTQGGG 821
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q7UCZ1 PRELIMINARY; PRT; 151 AA.
AC Q7UCZ1;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016981; AAP16542.1; -.
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

```



```
Query Match 14.3%; Score 112; DB 16; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.11;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETLQNGFR----- 98
DB 17 GIAAAGYDLANSEYNFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLAVVAQGS 76

QY 99 -NNATIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 SNRAKIDQTDYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 12
Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curilin subunit precursor.
DE CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.P., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Moxley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 14.3%; Score 112; DB 16; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.12;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETLQNGFR----- 98
DB 26 GIAAAGYDLANSEYNFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLTVVAQGS 85

QY 99 -NNATIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 86 SNRAKIDQTDYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138

RESULT 13
Q83R7 PRELIMINARY; PRT; 160 AA.
AC Q83R7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curilin subunit precursor, similar to CsgA.
DE CSGB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
```

```
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015131; AAN42658.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 14.3%; Score 112; DB 16; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.12;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETLQNGFR----- 98
DB 26 GIAAAGYDLANSEYNFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLAVVAQGS 85

QY 99 -NNATIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 86 SNRAKIDQTDYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138

RESULT 14
Q7X244 PRELIMINARY; PRT; 151 AA.
AC Q7X244;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
DE CSGB.
OC Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56671.1; -.
KW Complete proteome.
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.1%; Score 110; DB 2; Length 151;
Best Local Similarity 27.2%; Pred. No. 0.16;
Matches 31; Conservative 22; Mismatches 47; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGQADNSTIETLQNGFR----- 98
DB 17 GIASATSYDLAHSEYNFAVNELSKSFNQAAIIGQVTNNSAKMEQEGSKLLSVVSEQG 76

QY 99 -NNATIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 SNRAKVDSQAYNF-AYIAQSGHSNDASISQSNYGNNTAMIIQKSGNKANITQY 129

RESULT 15
Q8PD38 PRELIMINARY; PRT; 1333 AA.
AC Q8PD38;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Ice nucleation protein.
DE XCC0507.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=20202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Canavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE012148; AAM39823.1; -.
DR GO; GO:009279; C:external outer membrane (sensu Gram-negativ. . .; IEA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 68.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 40.
KW Complete proteome.
SQ SEQUENCE 1333 AA; 131306 MW; 326078458D0E4842 CRC64;

```

Query Match 13.9%; Score 108.5; DB 16; Length 1333;
Best Local Similarity 26.3%; Fred. No. 3;
Matches 46; Conservative 22; Mismatches 48; Indels 59; Gaps 9;

```

QY 5 KVAFAAIIWVG-----SALAGVWPQGGGNGGNSGPDSTLSIYQYGS 53
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 856 QTAGYKSLITGYGSTQTAQSSSLIAGY-----GSSWAGPDSL-IAGYST 903
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 54 NAALYDQLVTRVVTHEMAHAGYNGADVQGDNSTIELTQNGFRNNATID----- 104
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 904 QTAGYDSFLT-----AGYGS-----TQAQSSSLITGYGSTSTAFQSSLIAGYS 950
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 105 -QWNAKNSDITVGOYGGNNAL-----VNOTASDSSVMVQVQFGNNATA 148
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 951 TQTAGYESTLTAG-YGSTQTAEISWLTGYGSTQTAGHGSILT--AGYGSNSTA 1002
   : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: August 2, 2004, 14:54:41
Job time : 30.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAAPRAIVVSGSALA.....DSSVMVRQVFGNNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 775 | 100.0 | 151 | 3 | AAB36354 |
| 2 | 712 | 91.9 | 151 | 3 | AAB36351 |
| 3 | 683 | 88.1 | 151 | 2 | AAR74625 |
| 4 | 683 | 88.1 | 151 | 3 | AAB36341 |
| 5 | 678 | 87.5 | 151 | 2 | AAW23570 |
| 6 | 662 | 85.4 | 151 | 3 | AAB36350 |
| 7 | 617 | 79.6 | 151 | 3 | AAB36353 |
| 8 | 613 | 79.1 | 151 | 3 | AAB36355 |
| 9 | 605 | 78.1 | 151 | 3 | AAB36346 |
| 10 | 603 | 77.8 | 151 | 3 | AAB36347 |
| 11 | 601 | 77.5 | 151 | 3 | AAB36349 |
| 12 | 600 | 77.4 | 151 | 3 | AAB36352 |
| 13 | 568 | 73.3 | 151 | 3 | AAB36348 |
| 14 | 509 | 65.7 | 151 | 3 | AAB36343 |
| 15 | 504 | 65.0 | 151 | 7 | ABR82651 |
| 16 | 498 | 64.3 | 120 | 2 | AAR62761 |
| 17 | 498 | 64.3 | 120 | 2 | AAW23569 |
| 18 | 435 | 55.1 | 142 | 2 | AAR52664 |
| 19 | 359 | 46.3 | 122 | 2 | AAR52663 |
| 20 | 147 | 19.0 | 45 | 3 | AAB36316 |
| 21 | 132 | 17.0 | 22 | 3 | AAB36318 |
| 22 | 113 | 14.6 | 24 | 7 | ABR82644 |
| 23 | 111 | 14.3 | 22 | 3 | AAB36322 |
| 24 | 111 | 14.3 | 22 | 3 | AAB36327 |
| 25 | 111 | 14.3 | 22 | 3 | AAB36337 |

ALIGNMENTS

RESULT 1
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
AC AAB36354;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UTVI-); UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64630.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

26 109 14.1 23 3 AAB36340
27 109 14.1 23 3 AAB36324
28 109 14.1 23 3 AAB36319
29 102 13.2 26 7 ABR82649
30 96 12.4 19 3 AAB36323
31 96 12.4 19 3 AAB36336
32 96 12.4 19 3 AAB36328
33 94.5 12.2 151 3 AAB36342
34 92 11.9 1074 6 ABU22692
35 91.5 11.8 502 7 ABR82642
36 90 11.6 24 7 ABR82647
37 89.5 11.5 151 3 AAB36344
38 89 11.5 24 7 ABR82647
39 89 11.5 354 7 ABQ23520
40 87.5 11.3 249 3 AAY69523
41 87.5 11.3 447 3 AAG29728
42 87.5 11.3 468 3 AAG29727
43 87.5 11.3 842 5 ABF66189
44 87.5 11.3 1419 5 ABP69842
45 87.5 11.3 1477 5 ABP69841

AAB36340 Salmonell
AAB36324 Salmonell
AAB36319 Salmonell
ABR82649 E. coli V
AAB36323 Salmonell
AAB36336 Salmonell
AAB36328 Salmonell
AAB36342 Salmonell
ABU22692 Protein e
ABU22692 Leishmani
ABR82642 E. coli N
ABR82647 E. coli c
ABR82647 E. coli c
ABO23520 Mycobacte
AAY69523 Anti-CD38
AAG29728 Arabidops
AAG29727 Arabidops
ABF66189 Bifidobac
ABP69842 Human pol
ABP69841 Human pol

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 2.5e-67;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDIIVGQYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDIIVGQYGG 120
 QY 121 NNAALVNQTSASSVWVQVFGNNATANQY 151
 DB 121 NNAALVNQTSASSVWVQVFGNNATANQY 151

RESULT 2
 AAB36351
 ID AAB36351 standard; protein; 151 AA.
 AC AAB36351;
 DT 26-FEB-2001 (first entry)
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 PN WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 XX N-PSDB; AAC64627.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 91.9%; Score 712; DB 3; Length 151;
 Best Local Similarity 91.1%; Pred. No. 3.3e-61;
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDI 113
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDI 113
 QY 114 TVGQYGGNNAALVNQTSASSVWVQVFGNNATANQY 151
 DB 114 TVGQYGGNNAALVNQTSASSVWVQVFGNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 DE Agfa sequence.
 DE Salmonella; Agfa; vaccine.
 KW Salmonella.
 OS Salmonella.
 XX WO9425598-A2.
 XX 10-NOV-1994.
 XX 26-APR-1994; 94WO-IB000207.
 XX 26-APR-1993; 93US-00054452.
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING/) KING J.
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
 XX WPI: 1994-358275/44.
 DR N-PSDB; AAQ87467.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The Salmonella AgfA protein and DNA are used in vaccine and genetic
 CC immunization compositions, respectively, to elicit an immune response to
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated
 CC on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 151 AA;
 Query Match 88.1%; Score 683; DB 2; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 4
 AAB36341
 ID AAB36341 standard; protein; 151 AA.
 AC AAB36341;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
 XX
 KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 DR WPI: 2000-572631/55.
 DR N-PSDB; AAC64617.
 XX
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 PS
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 88.1%; Score 683; DB 3; Length 151;
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKSDITVGYGG 120
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKSDITVGYGG 120
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 RESULT 5
 AAW23570
 ID AAW23570 standard; protein; 151 AA.
 AC AAW23570;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfA.
 XX
 KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 123
 FT /note= "Encoded by GCC"
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;

```

XX WPI; 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteriaceae family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfA encoded by the full agfA gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteriaceae. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridize to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
XX
XX Query Match 87.5%; Score 678; DB 2; Length 151;
XX Best Local Similarity 89.4%; Pred. No. 6.6e-58;
XX Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADYDQVLTVRVVTHEMAHARRNATIDQWNAKNSDITVGGYGG 120
XX DB 61 SDARKSETTITQSGYNGADYDQVLTVRVVTHEMAHARRNATIDQWNAKNSDITVGGYGG 120
XX
XX QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
XX DB 121 NNPAALVNTASDSSVMVRQVGFNNATANQY 151
XX
XX RESULT 6
XX AAB36350
XX ID AAB36350 standard; protein; 151 AA.
XX AC
XX XX AAB36350;
XX
XX DT 26-FEB-2001 (first entry)
XX DE
XX XX AgfA::PT3#5 amino acid sequence SEQ ID NO:20.
XX KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX KW vaccine; immune response; immunogen.
XX
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64626.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant agfA

```

```

PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant AgfA
XX protein containing a replacement segment or segments of a recombinant AgfA
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant AgfA protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;

```

```

Query Match 85.4%; Score 662; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.4e-56;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADYDQVLTVRVVTHEMAFA-----FRNATIDQ 105
XX DB 61 SDARK-----YDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNATIDQ 105
XX
XX QY 106 WNAKNSDITVGGYNGAALVNQTSDDSSVMVRQVGFNNATANQY 151
XX DB 106 WNAKNSDITVGGYNGAALVNQTSDDSSVMVRQVGFNNATANQY 151

```

```

RESULT 7
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
XX AC AAB36353;
XX
XX DT 26-FEB-2001 (first entry)
XX DE
XX XX AgfA::PT3#8 amino acid sequence SEQ ID NO:26.
XX KW Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;
XX KW vaccine; immune response; immunogen.
XX
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX
XX

```

```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 79.6%; Score 617; DB 3; Length 151;
XX Best Local Similarity 73.6%; Pred. No. 5.7e-52;
XX Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSSGDPSTLSIYQVGSANAALQ 60
XX Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSSGDPSTLSIYQVGSANAAL--- 57
XX
XX QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----F 97
XX Db -----YDQLVTRVVTHEMAHAGYGADVGQGDNSTIELTQNGF 97
XX
XX QY 98 RNNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
XX Db 98 RNNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
XX
XX RESULT 8
XX ID AAB36355 standard; protein; 151 AA.
XX AC AAB36355;
XX DT 26-FEB-2001 (first entry)
XX XX Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO2000060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64631.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX Disclosure; Page 139; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 79.1%; Score 613; DB 3; Length 151;
XX Best Local Similarity 74.6%; Pred. No. 1.4e-51;
XX Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSSGDPSTLSIYQVGSANAALQ 60
XX Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSSGDPSTLSIYQVGSANAALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAHAFR 98
XX Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHA-- 118
XX
XX QY 99 RNNATIDOWNAKNSDITVQYGGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
XX Db 119 -----GGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
XX
XX RESULT 9

```

AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PI
 XX WPI; 2000-672631/65.
 DR
 DR N-PSDB; AAC64622.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 135; 139pp; English.
 PS
 XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 151 AA;
 SQ
 Query Match 78.1%; Score 605; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 8.3e-51;
 Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;
 Qy 1 MKLLKVAFAAIVVSGSALAGVWPQWGGGNGHNGGNSGPDSTLSIYQVGSNAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVWPQWGGGNGHNGGNSGPDSTLSIYQVGSNAALALQ 60

Qy 61 SDARKSETTITQSGYNGADYDQLVTVVVTTHMAHAFRNATIDONAKNSDITVGOYGG 120
 DB ||||||||||||||||||||| : |||||||||||||||||||||||||
 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDONAKNSDITVGOYGG 120
 Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
 DB ||||||| : : |||||||
 121 NNAALVNYDQLVTVVVTTHMAHANATANQY 151

RESULT 10

AAB36347
 ID AAB36347 standard; protein; 151 AA.
 XX

AC AAB36347;
 XX

DT 26-FEB-2001 (first entry)
 XX

DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
 XX

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
 KW vaccine; immune response; immunogen.
 XX

OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX

XX WO200060102-A2.
 XX

XX 12-OCT-2000.
 XX

XX 05-APR-2000; 2000WO-CA000356.
 PF

XX 05-APR-1999; 99US-0127888P.
 PR

XX (UYVI-) UNIV VICTORIA.
 XX

XX White AP, Doran JL, Collison SK, Kay WW;
 PI

XX WPI; 2000-672631/65.
 DR

XX N-PSDB; AAC64623.
 DR

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX

XX Disclosure; Page 136; 139pp; English.
 XX

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention
XX Sequence 151 AA;
SQ

Query Match 77.8%; Score 603; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.3e-50;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 LVRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 151

RESULT 11
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64625.
XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.
PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF1/7/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal. CC comprising separating an amino acid polymer comprising a recombinant AgfA protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant AgfA protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 77.5%; Score 601; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 2e-50; Mismatches 25; Indels 0; Gaps 0;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
AAB36352
ID AAB36352 standard; protein; 151 AA.
XX
AC AAB36352;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#7 amino acid sequence SEQ ID NO:24.
XX
KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX
DR N-PSDB; AAC64628.
XX

Recombinant agfA gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant AgfA protein useful for eliciting immune response in animal.
PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfA gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 2.5e-50;
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWAKNSDITVGYGG 120
 Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWAKNSDITVTRVVT 120
 QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
 Db 121 HEMAHANQTASDSSVWVRQVGFNNATANQY 151

RESULT 13

AAAB36348
 ID AAB36348 standard; protein; 151 AA.

XX AAB36348;

XX 26-FEB-2001 (first entry)

DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.3%; Score 568; DB 3; Length 151;
 Best Local Similarity 80.1%; Pred. No. 3.3e-47;
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWAKNSDITVGYGG 120

Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151

RESULT 14

AAAB36343
 ID AAB36343 standard; protein; 151 AA.

XX AAB36343;

XX 26-FEB-2001 (first entry)

XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; immunogen.

XX Escherichia coli.

XX WO200060102-A2.

XX 12-OCT-2000.

XX PD

PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64619.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 135; 139pp; English.
 PS
 XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SF17/TA) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 65.7%; Score 509; DB 3; Length 151;
 Best Local Similarity 68.2%; Pred. No. 1.8e-41;
 Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAIAIVFSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
 Db 61 TDARNSDLTITQGGGNGADVGQSDSSIDLTRQFGNSATLDOWNKNSMTVKQFGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 Db 121 GNGAAVDQTASNSNVTVQVGFNNATAHQY 151
 RESULT 15
 ABR82651
 ID ABR82651 standard; protein; 151 AA.
 XX
 AC ABR82651;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE E. coli CsgA subunit 15 kDa protein.
 XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
 XX Escherichia coli.
 XX WO2003064446-A2.
 XX 07-AUG-2003.
 XX 30-JAN-2003; 2003WO-EP000943.
 XX 31-JAN-2002; 2002GB-00002275.
 PR (HANS-) HANSA MEDICAL RES AB.
 PA Bjoerck L, Olsen A, Wikstroem M, Herwald H;
 XX WPI; 2003-646136/61.
 DR N-PSDB; ACF36153.
 XX New isolated peptide capable of binding a mammalian plasma protein,
 PT useful in the manufacture of a medicament for the prevention and/or
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
 PT or Shigella infections.
 XX Disclosure; Page 41-42; 42pp; English.
 PS The invention relates to an isolated peptide capable of binding a
 XX mammalian plasma protein or of generating an immune response in a mammal
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
 CC antibody is useful for treating a bacterial infection in a human or
 CC animal or in the manufacture of a medicament for the prophylactic
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
 CC or Shigella infection. The peptide that is immobilized on a solid support
 CC is also useful as a reagent for determining the ability of a plasma
 CC protein to bind to bacteria. The present sequence represents an E. coli
 CC 15 kDa protein
 XX
 SQ Sequence 151 AA;
 Query Match 65.0%; Score 504; DB 7; Length 151;
 Best Local Similarity 67.5%; Pred. No. 5.5e-41;
 Matches 102; Conservative 15; Mismatches 34; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAIAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 Db 1 MKLLKVAFAIAIVFSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
 Db 61 TDARNSDLTITQGGGNGADVGQSDSSIDLTRQFGNSATLDOWNKNSMTVKQFGG 120
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
 Db 121 GNGAAVDQTASNSNVTVQVGFNNATAHQY 151
 Search completed: August 2, 2004, 14:48:29
 Job time : 45.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------------------|-------------------|
| 1 | 678 | 87.5 | 151 | 1 US-08-233-788A-59 | Sequence 59, Appl |
| 2 | 498 | 64.3 | 120 | 1 US-08-233-788A-57 | Sequence 57, Appl |
| 3 | 85 | 11.0 | 975 | 4 US-09-328-352-4764 | Sequence 4764, Ap |
| 4 | 84.5 | 10.9 | 738 | 3 US-08-864-038A-3 | Sequence 3, Appl1 |
| 5 | 84 | 10.8 | 943 | 4 US-09-056-556-204 | Sequence 204, App |
| 6 | 84 | 10.8 | 943 | 4 US-09-072-596-199 | Sequence 199, App |
| 7 | 84 | 10.8 | 943 | 4 US-09-477-135A-131 | Sequence 131, App |
| 8 | 84 | 10.8 | 943 | 4 US-09-072-967-204 | Sequence 204, App |
| 9 | 82 | 10.6 | 2123 | 3 US-08-968-685A-10 | Sequence 10, Appl |
| 10 | 81.5 | 10.5 | 558 | 4 US-09-252-991A-30983 | Sequence 30983, A |
| 11 | 80 | 10.3 | 339 | 4 US-09-252-991A-32096 | Sequence 32096, A |
| 12 | 80 | 10.3 | 1207 | 4 US-08-489-039A-11518 | Sequence 11518, A |
| 13 | 79.5 | 10.3 | 745 | 4 US-08-336-115C-6 | Sequence 6, Appl1 |
| 14 | 79.5 | 10.3 | 1415 | 4 US-09-252-991A-26438 | Sequence 26438, A |
| 15 | 78.5 | 10.1 | 714 | 4 US-09-841-786-4 | Sequence 4, Appl1 |
| 16 | 78.5 | 10.1 | 3241 | 4 US-08-841-786-1 | Sequence 1, Appl1 |
| 17 | 78 | 10.1 | 892 | 4 US-08-336-447A-5 | Sequence 5, Appl1 |
| 18 | 78 | 10.1 | 918 | 4 US-09-200-650E-1 | Sequence 1, Appl1 |
| 19 | 77.5 | 10.0 | 415 | 4 US-09-025-769B-280 | Sequence 280, App |
| 20 | 77.5 | 10.0 | 518 | 3 US-09-043-123-2 | Sequence 2, Appl1 |
| 21 | 77 | 9.9 | 1160 | 3 US-08-808-599A-24 | Sequence 24, Appl |
| 22 | 76.5 | 9.9 | 873 | 4 US-09-336-447A-13 | Sequence 13, Appl |
| 23 | 75.5 | 9.7 | 211 | 1 US-08-276-852-34 | Sequence 34, Appl |
| 24 | 75.5 | 9.7 | 211 | 1 US-08-133-011-16 | Sequence 16, Appl |
| 25 | 75.5 | 9.7 | 211 | 1 US-08-322-730A-16 | Sequence 16, Appl |
| 26 | 75.5 | 9.7 | 211 | 1 US-08-387-874-16 | Sequence 16, Appl |
| 27 | 75.5 | 9.7 | 211 | 1 US-08-899-575-34 | Sequence 34, Appl |

| | | | | | |
|----|------|-----|-----|---------------------|-------------------|
| 28 | 75.5 | 9.7 | 211 | 1 US-08-899-575-34 | Sequence 34, Appl |
| 29 | 75.5 | 9.7 | 211 | 2 US-08-383-619-16 | Sequence 16, Appl |
| 30 | 75.5 | 9.7 | 211 | 3 US-08-907-739-16 | Sequence 16, Appl |
| 31 | 75.5 | 9.7 | 211 | 4 US-09-729-597-16 | Sequence 16, Appl |
| 32 | 75.5 | 9.7 | 211 | 5 PCT-US93-08364-16 | Sequence 16, Appl |
| 33 | 75.5 | 9.7 | 211 | 5 PCT-US95-08743-34 | Sequence 34, Appl |
| 34 | 75.5 | 9.7 | 238 | 4 US-09-495-880A-42 | Sequence 42, Appl |
| 35 | 75.5 | 9.7 | 244 | 2 US-08-553-497A-22 | Sequence 22, Appl |
| 36 | 75.5 | 9.7 | 266 | 4 US-09-495-880A-26 | Sequence 26, Appl |
| 37 | 75.5 | 9.7 | 293 | 3 US-08-438-745-6 | Sequence 6, Appl1 |
| 38 | 75.5 | 9.7 | 293 | 3 US-08-438-745-6 | Sequence 6, Appl1 |
| 39 | 75.5 | 9.7 | 293 | 3 US-09-219-019-4 | Sequence 4, Appl1 |
| 40 | 75.5 | 9.7 | 293 | 3 US-09-219-019-6 | Sequence 6, Appl1 |
| 41 | 75.5 | 9.7 | 293 | 5 PCT-US94-05669A-4 | Sequence 4, Appl1 |
| 42 | 75.5 | 9.7 | 293 | 5 PCT-US94-05669A-6 | Sequence 6, Appl1 |
| 43 | 75.5 | 9.7 | 432 | 4 US-09-403-089A-1 | Sequence 1, Appl |
| 44 | 75.5 | 9.7 | 461 | 2 US-08-463-587A-26 | Sequence 26, Appl |
| 45 | 75.5 | 9.7 | 461 | 2 US-08-463-667A-4 | Sequence 4, Appl1 |

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 692-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 87.5%; Score 678; DB 1; Length 151;
Best Local Similarity 89.4%; Pred No. 48-61;
Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNHGNGSGPDSTLSTIYQVGSANAALALQ 60
|||||

Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTSLIYQVGSAAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYDQADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120
QY 121 NPAALVNTASDSSVMVQVFGNATANQY 151
Db 121 NPAALVNTASDSSVMVQVFGNATANQY 151

RESULT 2

US-08-233-788A-57

; Sequence 57, Application US/08233788A

; Patent No. 5635617

; GENERAL INFORMATION:

; APPLICANT: Doran, James L.

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, Karen S.

; APPLICANT: Clouhier, Sharon C.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

; TITLE OF INVENTION: OF SALMONELLA

; NUMBER OF SEQUENCES: 61

; CORRESPONDENCE ADDRESS:

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/233,788A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 57:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 120 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-233-788A-57

Query Match

; Best Local Similarity 64.3%; Score 498; DB 1; Length 120;

; Mismatches 97; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGGNGGNSGPDSTLSIYQVGSAAALALQSDARKSETTITQSGYNGADY 81

Db 1 VVPQWGGGNGGNGGNSGPDSTLSIYQVGSAAALALQSDARKSETTITQSGYNGADY 60

QY 82 DQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQVGGNNAALVNTASDS 133

Db 61 QGGADNSTIELTQNGFRNATIDQWNAKNSDITVQVGGNNAALVNTASDS 112

RESULT 3

US-09-328-352-4764

; Sequence 4764, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4764
; LENGTH: 975
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4764

Query Match 11.0%; Score 85; DB 4; Length 975;

Best Local Similarity 25.3%; Pred No. 4.2;

Matches 38; Conservative 18; Mismatches 52; Indels 42; Gaps 7;

QY 15 SGSALAGVVPQWGGGNGGNGGNSGPDSTLSIYQYG-----SANA 55

Db 300 AGNIA-----SGNHEHYGIGNGGDDVDITAPITGVLTNIGNSFTLIGNSSSSSVNT 353

QY 56 ALALQSDARKSETTI-----TQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNA 108

Db 354 APTTSTNTVNDNTIDNGSGGTGSGNGSG-DGLLNGAASNGEH-----NYGIGNG 408

QY 109 KNSDIT-----VQYCGNNAALVNTASDS 133

Db 409 DQVDITSPITGIFNFGNSFSLIGNSSSSS 438

RESULT 4

US-08-864-038A-3

; Sequence 3, Application US/08864038A

; Patent No. 6001592

; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR

; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID

; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING

; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSER: 812-5 Hirano

; STREET: Isehinden

; CITY: Tsu-city

; STATE: Mie-prefecture

; COUNTRY: JAPAN

; ZIP: 514-01

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Microsoft Windows 95

; SOFTWARE: Word Perfect 6.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A

; FILING DATE: May 28, 1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 8-184459

; FILING DATE: 15-July-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: C. Bruce Hamburg

; REGISTRATION NUMBER: 22,389

; REFERENCE/DOCKET NUMBER: F-5610

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212)986-2340

; TELEFAX: (212)953-7733

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 738

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

```

; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3

Query Match 10.9%; Score 84.5; DB 3; Length 738;
Best Local Similarity 35.5%; Pred. No. 3.3;
Matches 27; Conservative 4; Mismatches 22; Indels 23; Gaps 3;

QY 3 LKVAAPAAIVVSSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQSD 62
Db 419 LKSSASASASASASAG-----GGGGGNGGNGGGG-----GGAGALA----- 460

QY 63 ARKSETTITQSGYNG 78
Db 461 -----ALAAAGAGGG 471

RESULT 5
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-204

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSGPDSTLSIYQGSANAALALQSDARKSETTITQS 73
Db 464 GSGNIGVFNVGSGSLGNYNGSGN-----LGIYNGFGNVG-----DYNV 503

QY 74 GYNGADYDQLVTRVVTHEMAFAFNATIDOWNAKNSDITVGYQYGGN----- 121

RESULT 6
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-199

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSGPDSTLSIYQGSANAALALQSDARKSETTITQS 73
Db 464 GSGNIGVFNVGSGSLGNYNGSGN-----LGIYNGFGNVG-----DYNV 503

QY 74 GYNGADYDQLVTRVVTHEMAFAFNATIDOWNAKNSDITVGYQYGGN----- 121
Db 504 GFGNAGDFNQ-----GFANTGNNIGFANTGNNIGLSDNQGGFNIAAGWNS 553

QY 122 ---NAALVNTASDSSVM---VRQVGFGNATAN 149
Db 554 GTGNSGLFNSGNNVGVIFNAGTGVNGVANS GTGN 587

RESULT 7
US-09-477-135A-131
```

```
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Nanco, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 5288
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; US-09-477-135A-131

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGDPSTLSIYQYGSANAALALQSDARKSETTITQS 73
Db 571 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFGNVG-----DYNV 610

QY 74 GYNGADYDQLVTRVVTTHMAHAFENNAATIDQWNAKNSDITVQYGGN-----121
Db 611 GFENAGDFNQ-----GFANTGNNNIGFANTGNNNIGLIGLSDNQOQGFNIASGWN 660

QY 122 ---NAALVNOTASDSSVM---VRQVGFNGNATAN 149
Db 661 GTGNSGLFNSGTNNVGFNAGTGNVGIANSCTGN 694

RESULT 8
US-09-072-967-204
; Sequence 204, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:

US-09-072-967-204
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-09-072-967-204

Query Match 10.8%; Score 84; DB 4; Length 943;
Best Local Similarity 24.7%; Pred. No. 5.1;
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPQWGGG-GNHN-GGNSSGDPSTLSIYQYGSANAALALQSDARKSETTITQS 73
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFGNVG-----DYNV 503

QY 74 GYNGADYDQLVTRVVTTHMAHAFENNAATIDQWNAKNSDITVQYGGN-----121
Db 504 GFENAGDFNQ-----GFANTGNNNIGFANTGNNNIGLIGLSDNQOQGFNIASGWN 553

QY 122 ---NAALVNOTASDSSVM---VRQVGFNGNATAN 149
Db 554 GTGNSGLFNSGTNNVGFNAGTGNVGIANSCTGN 587

RESULT 9
US-08-968-685A-10
; Sequence 10, Application US/08968685A
; Patent No. 6214981
; GENERAL INFORMATION:
; APPLICANT: TUCKER, KENNETH
; APPLICANT: PLOSILA, LAURA
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USSES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2123 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
```


US-08-968-685A-10

```

Query Match      10.6%; Score 82; DB 3; Length 2123;
Best Local Similarity 25.8%; Pred. No. 24;
Matches 31; Conservative 17; Mismatches 38; Indels 34; Gaps 5

QY 1 MKLLKVAFAAIV-----SGSALAGVWPQGGGNNHNGGNSGGPDSITUYQYGSANAA 56
      : : : : :
Db 44 LSFARIALAVLIGATLNGSAYAGIGISEADGG--KGGANARG-DKSAI-----GDIAQA 97
      : : : : :

QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTENAHAFRNNAITDOWNAKNSDITVG 116
      : : : : :
Db 98 LGSQSGIA-----IGSNKIYVHNSNNNAIGAKSNGESIAIG 133

```

```

RESULT 10
US-09-252-991A-30983
; Sequence 30983, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCES: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30983
; LENGTH: 558
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30983

```

| | | | | |
|-----------------------|--------|--|----------------|-------------------|
| Query Match | 10.5% | Score 81.5; | DB 4; | Length 558; |
| Best Local Similarity | 30.7%; | Pred.No. 4.5; | | |
| Matches | 27; | Conservative 11; | Mismatches 39; | Indels 11; Gaps 2 |
| QY | 16 | GSALGV-----VPGWGGGHHNGGNSGPDSTLSIYQVSANAALAQSPARKSETTI | 70 | |
| Db | 220 | GAAFAGVGEALVGLGGHLDGGADG-----AVHEGEHFAQVLRSDDEEAGGAVE | 273 | |
| QY | 71 | TQSGYNGADYDQLVTRVVTTHMAHAFR | 98 | |
| Db | 274 | VHHAGRGGLDPHLLLORAAATHAVARER | 301 | |

RESULT 11
US-09-252-991A-32096
; Sequence 32096, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32096
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32096

Query Match 10.3%; Score 80; DB 4; Length 339;

| | | | | | | | | | |
|----|-----------------------|---|------------------------------|---------|------------------|----------------|------------|----------|----|
| | Best Local Similarity | 20.1%; Pred. No. | 3.3; | Matches | 31; Conservative | 34; Mismatches | 61; Indels | 28; Gaps | 6; |
| Qy | 5 | KVAFAAIVSGSALAGVWPQG- | -GNNHNGCGNSSPDSTLSIYQYSANAAL | 57 | : : | : | : | : | : |
| Dd | 58 | KVSNFGL--NNASVGSIKASGVGNVAAADNNQAANALAASADAASFVFGTATAS- | 114 | :: | ::: | : | : | : | : |
| Qy | 58 | ALOSDARKSETTTTOSGYNGA-DYDLTVRVVTHEVAHFNR--- | NATTIDOWNAKNSDI | 113 | | : | : | : | : |
| Dd | 115 | -----TSVLQSYGNTLNNYNPENTASLNSANNVSGNLGVNAAGFNQQKNLD | 164 | | : | : | : | : | : |
| Qy | 114 | TV---GOYGENNAALVNQTASDSVMVRQVGF | 143 | | : | : | : | : | : |
| Dd | 165 | AAAVSNCOYSTAGSAASOTSOTSGTNVTNSANYAYG | 198 | | : | : | : | : | : |

RESULT 12
US-09-489-039A-11518
; Sequence 11518, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11518
; LENGTH: 1207
; TYPE: prt
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11518

| | Query Match | 10.3% | Score 80; | DB 4; | Length 1207; |
|----|-----------------------|---|----------------|------------|--------------|
| | Best Local Similarity | 23.5%; | Pred. No. 18; | | |
| | Matches 36; | Conservative 23; | Mismatches 66; | Indels 28; | Gaps 5; |
| QY | 21 | GVVPQWGGGGNHN--GGGNSSGPDSTLSIYQVS-----ANAALALQSDARKSETTITQS | 73 | | |
| Db | 482 | GPWQSQPAYNNHQFSGAANGAESMISTFSLGTLASGVSSQAALQSKSEAFQSTLGR | 541 | | |
| QY | 74 | GYGNGADYDQLVTRV-----VTHEMAHAFRNNTATDQWNAKNSDITVGQ | 117 | | |
| Db | 542 | GFSDGVSQDQAYSELNIGENVSSQNTASQLINQQAQKFMDFQVDD---SHSDAVKGA | 598 | | |
| QY | 118 | YG--GNNALVNQTASDSSVMVRQVQFGNNATA | 148 | | |
| Db | 599 | FAMCATGTTLDVDOAAWMLMPMVGKAAAKAA | 631 | | |

```

RESULT 13
US-09-336-115C-6
; Sequence 6; Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,259
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori

```

```
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)....(20)
;
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match          10.3%; Score 79.5; DB 4; Length 745;
Best Local Similarity 20.5%; Pred. No. 11;
Matches 35; Conservative 27; Mismatches 70; Indels 39; Gaps 6;

QY 8 AFAIIVSGSALAGVVPQW-----GGGNHNGGN----- 37
Db 102 AYQAVFLAINAAVGL--WNTIGYAVCMGNGTSGPGSVIFNDQPGQDSTQITCNRFE 158
QY 38 SSGPDSTLSIYQYGSANALALQSDARKSETITQSGYNGADYDQLVTRVVTHEMAHAF 97
Db 159 STPGKSMISIDBFKXLEAYQIIQALKXQSGFPPLG-GNCTK----VSVNRYNECRQTA 213
QY 98 RNNATIDOW-NAKNSDITVGQYGGNNAALVNCTASDSSVMVRQVGFGNAT 147
Db 214 DINGGVYQFCRAKNGSSSSNGSGSTQTATTITQDGVTTITTYNNKAT 264

RESULT 14
US-09-252-991A-26438
; Sequence 26438, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26438
; LENGTH: 1415
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26438

Query Match          10.3%; Score 79.5; DB 4; Length 1415;
Best Local Similarity 28.9%; Pred. No. 25;
Matches 44; Conservative 17; Mismatches 74; Indels 17; Gaps 9;

QY 10 AAIIVSGSALAGVVPQWGGG--NHNGG--NSSGPDSTLSIYQYGSANALALQSDA-- 63
Db 930 ADFAISGQ-LKDHASHYGAGGLVGRNGLIRSSGQSTLSLGHGMNGLGLVGYSSAGG 988
QY 64 -RKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVGQYGN- 121
Db 989 LADVASVDVSGNGQGLYGLGLGNVNSGTAHATASGKV----RGTDAAELGGLIGNL 1044
QY 122 NAALVNQTA-SDSSVMV-RQVG--FGNNATAN 149
Db 1045 NAAINNASHGDVSLQAGRYLGLIGHNQAGN 1076

RESULT 15
US-09-841-786-4
; Sequence 4, Application US/09841786
; Patent No. 6669940
; GENERAL INFORMATION:
; APPLICANT: NAGARAJA, T. G.
; APPLICANT: STEWART, GEORGE C.
; APPLICANT: NARAYANAN, SANJEEV K.
```

```
; APPLICANT: CHENGAPPA, M. M.
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN
; TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF
; FILE REFERENCE: 30296
; CURRENT APPLICATION NUMBER: US/09/841,786
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 09/558,257
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
US-09-841-786-4

Query Match          10.1%; Score 78.5; DB 4; Length 714;
Best Local Similarity 22.8%; Pred. No. 13;
Matches 31; Conservative 16; Mismatches 56; Indels 33; Gaps 6;

QY 12 IVVSGSALAGVVPQWGGG-----NHNGGN-----SSGPDSTLSIYQYGSANALALQSDAR 64
Db 35 VTSSDSTFVGA---WGGSAALQWNHIGSGNSNISAGLAGAAAANNIOKTSALVVKNSDIR 91
QY 65 KSE-----TTITQSGYNGADY---DQLVTRVVTHEMAHAFRNNAT 102
Db 92 NANKFKNNALSGTQVAAGAGLEAVKESG-GQCKSYLLGTSASINLVNNEVSAKSENNTV 150
QY 103 IDQWNAKNSDITVGQY 118
Db 151 AGESESQKMDVDVTAY 166

Search completed: August 2, 2004, 14:58:36
Job time : 12 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)

1287.123 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANCY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | DB ID | Description |
|------------|-------------|-------|--------|-------|----------------------|
| 1 | 506 | 65.3 | 151 | 12 | US-09-741-873B-4 |
| 2 | 506 | 65.3 | 151 | 12 | US-09-741-873B-4 |
| 3 | 428 | 55.2 | 131 | 12 | US-09-741-873B-2 |
| 4 | 428 | 55.2 | 131 | 12 | US-09-741-873B-2 |
| 5 | 95 | 12.3 | 400 | 16 | US-10-437-963-186417 |
| 6 | 92 | 11.9 | 1074 | 12 | US-10-282-122A-50616 |
| 7 | 89 | 11.5 | 354 | 9 | US-09-820-843A-21 |
| 8 | 87.5 | 11.3 | 249 | 9 | US-09-730-374-3 |
| 9 | 87.5 | 11.3 | 249 | 16 | US-10-704-206-3 |
| 10 | 87.5 | 11.3 | 445 | 15 | US-10-369-493-20638 |
| 11 | 87.5 | 11.3 | 1448 | 9 | US-10-408-765A-998 |
| 12 | 87 | 11.2 | 597 | 6 | US-09-793-306-146 |
| 13 | 86 | 11.1 | 974 | 12 | US-10-282-122A-49999 |
| 14 | 85.5 | 11.0 | 2204 | 10 | US-10-282-122A-64364 |
| 15 | 85 | 11.0 | 254 | 10 | US-09-880-748-905 |

Sequence 905, App
Sequence 54647, A
Sequence 54379, A
Sequence 131, App
Sequence 131, App
Sequence 131, App
Sequence 199, App
Sequence 204, App
Sequence 64369, A
Sequence 162284,
Sequence 283675,
Sequence 72467, A
Sequence 50037, A
Sequence 147748,
Sequence 32, Appl
Sequence 33, Appl
Sequence 4, Appl
Sequence 48412, A
Sequence 56041, A
Sequence 275468,
Sequence 57763, A
Sequence 9, Appl
Sequence 54094, A
Sequence 4894, Ap
Sequence 246651,
Sequence 189881,
Sequence 9134, Ap
Sequence 187064,
Sequence 53269, A
Sequence 1847, Ap

254 12 US-10-293-418-905
928 12 US-10-282-122A-54647
978 12 US-10-282-122A-54379
943 9 US-09-995-634-131
943 10 US-09-997-182-131
943 10 US-09-997-181-131
943 14 US-10-193-002-199
943 14 US-10-084-843-204
3300 12 US-10-282-122A-64369
84 10.8 US-10-437-963-162284
84 10.8 US-10-282-122A-54094
83.5 10.8 US-10-437-963-162284
83.5 10.8 US-10-437-963-162284
288 12 US-10-437-963-162284
613 12 US-10-437-963-162284
793 12 US-10-282-122A-50037
145 16 US-10-437-963-147748
319 12 US-10-244-596-32
319 12 US-10-244-596-32
580 12 US-10-647-057-4
186 12 US-10-282-122A-49412
369 12 US-10-425-114-56041
486 12 US-10-424-599-275468
507 12 US-10-425-114-57763
2122 9 US-09-813-214A-9
485 12 US-10-282-122A-54094
224 9 US-09-738-626-4894
341 12 US-10-424-599-246651
703 16 US-10-437-963-108981
812 15 US-10-369-493-9134
342 16 US-10-437-963-187064
628 12 US-10-282-122A-53269
246 10 US-09-880-748-1847

ALIGNMENTS

RESULT 1

US-09-741-873B-4
Sequence 4, Application US/09741873B
Publication No. US20020081722A1
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873B
CURRENT FILING DATE: 2003-04-04
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 151
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-873B-4

Query Match 65.3%; Score 506; DB 12; Length 151;
Best Local Similarity 67.5%; Pred. No. 2.3e-42;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGSPDSTLSIYQVGSANAALAQ 60

[illegible]

RESULT. T 2

```

RECORD 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olesen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084

```

```

; CURRENT APPLICATION NUMBER: US709/741,873B
;
; CURRENT FILING DATE: 2003-04-04
;
; PRIOR APPLICATION NUMBER: SE 8801723-1
;
; PRIOR FILING DATE: 1998-05-06
;
; PRIOR APPLICATION NUMBER: US 08/978,878
;
; PRIOR FILING DATE: 1997-11-26
;
; PRIOR APPLICATION NUMBER: US 07/347,189
;
; PRIOR FILING DATE: 1989-05-04
;
; PRIOR APPLICATION NUMBER: US 07/789,437
;
; PRIOR FILING DATE: 1991-11-06
;
; PRIOR APPLICATION NUMBER: US 07/970,846
;
; PRIOR FILING DATE: 1992-11-03
;
; PRIOR APPLICATION NUMBER: US 08/187,865
;
; PRIOR FILING DATE: 1994-01-28
;
; PRIOR APPLICATION NUMBER: US 08/318,519
;
; PRIOR FILING DATE: 1994-10-05
;
; NUMBER OF SEQ ID NOS: 10
;
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 4
;
; LENGTH: 151
;
; TYPE: PRT
;
; ORGANISM: Escherichia coli
;
; US-09-741-873B-4
;

```

RESIST. 3

RES001.3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/197,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: FRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match          55.2%;   Score 428;   DB 12;   Length 131;
Best Local Similarity 64.1%;   Pred. No. 1.1e-34;
Matches 84;   Conservative 15;   Mismatches 32;   Indels 0;   Gaps 0;

Qy      21  GVVPQGGGNGHGGNGSSGPSTLSIYQYGSANAALAQSDARKSETTITQSGYNGAD 80
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      1  GVVPQYGGGNGHGGNGSSGPNSELNTYQYGGGNSALALQTDARNSDLTITCHGGGNGAD 60
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Qy      81  YDQLVTRVVTHMAHAFRNATIDWNAKNSDITVGOYGNNALVNOTASDSSVWVROV 140
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      61  VGGGSDSSSIDLTQRFGNSSALDQWNGKNSNTVKQFGGGNGAAVDQTASNSSVNVTOV 120
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

Qy      141  GFGNNATANQY 151
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      121  GFGNNATAHOY 131
        |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873B-2
Query Watch      55.2%   Score 428:   DB 12:   Length 131:

```

Best Local Similarity 64.1%; Pred. No. 1.1e-34;
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;

QY 21 GVVPQGGGNNHNGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPYQGGGNNHNGGNSGPNSELATYQYGGNSALA-QTDARNSDLIITQHGNGAD 60
QY 81 YDQIVRVVTHEMAHAFRNATIDQNAKNSDITVQYGGNNAALVNTASDSSVMVROV 140
Db 61 VGGGSDSSIDLQRFQFGNSATLDQNGKNSMTVRQFGGNGAAVDQTSNNSVNTQV 120
QY 141 GFGGNATANOY 151
Db 121 GFGGNATAHOY 131

RESULT 5

US-10-437-963-186417
; Sequence 186417, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 186417
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MPT4530_83218C.1.pap
US-10-437-963-186417

Query Match 12.3%; Score 95; DB 16; Length 400;
Best Local Similarity 23.6%; Pred. No. 0.52;
Matches 37; Conservative 20; Mismatches 62; Indels 38; Gaps 5;

QY 14 VSGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQ 73
Db 252 MNGSSSSSMFQLGGGGMSNGNS-----FSASMAAFWDEEAAMWATAVAPP 300
QY 74 GYNGADYDQIV-----TRVVTHEMAHAFRNATIDQNAKNSDIT 114
Db 301 GMGLAGMDVQAQQQQQVLDQAAGVAPTLVXHGGAATAGAMVVDX-SVMVVDVS 359
QY 115 VQYGGNNAALVNTASDSSVMVROVGF--GNNTAN 149
Db 360 SAVYGGATA-----TAFDLDMVESCGMFCGGGAGN 391

RESULT 6

US-10-282-122A-50616
; Sequence 50616, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 50616

; LENGTH: 1074

; TYPE: PRT

; ORGANISM: Burkholderia mallei

; US-10-282-122A-50616

Query Match 11.9%; Score 92; DB 12; Length 1074;

Best Local Similarity 28.1%; Pred. No. 3.5;

Matches 39; Conservative 21; Mismatches 69; Indels 10; Gaps 5;

QY 13 VVGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQ 72
Db 22 TASGULLA--APAWADTVSPGTDNVYGVDAIDPGVSTNQNTAYGAQAGAKVTGSYN 79
QY 73 SGYNGADYDQIVRVVTHEMAHAFRNATIDQNAKNS---DITVQYGGNNAALVNT 129
Db 80 IGYQAGQNVN-VIDIVSIGKQATASANDAIAIGTNTKASGPADIYMLNAGAG--STT 136
QY 130 ASDSSVM--VRQVGFNNA 146
Db 137 SPDGTVTLGIRNMLGESA 155

RESULT 7

US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:

; APPLICANT: Council of Scientific and Industrial Research

; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE

; FILE OF INVENTION: USEFUL AS ANTI-INFECTIVES

; FILE REFERENCE: Q63915

; CURRENT APPLICATION NUMBER: US/09/820,843A

; CURRENT FILING DATE: 2001-03-30

; NUMBER OF SEQ ID NOS: 118

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 21

; LENGTH: 354

; TYPE: PRT

; ORGANISM: M. tuberculosis

; FEATURE:

; NAME/KEY: misc_feature

```
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: GI1781260
US-09-820-843A-21

Query Match      11.5%; Score 89; DB 10; Length 354;
Best Local Similarity 24.1%; Pred. No. 1.7;
Matches 38; Conservative 17; Mismatches 63; Indels 40; Gaps 7;

QY 10 AAVVSGSALAG-VVPQWGGGNNHGG-----NSSGPDSTLSIYQGSAN 54
Db 200 AGDVNSGVNAGDVNTGLNSGNINTGFGNPGTLTGFSSAMTQAGFNS--GFENACTGN 257
QY 55 AALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDIT 114
Db 258 SGFGHNDPAGSGNGIQNSGFGNS-----GYVNTSTTSMFGNSGVLN 300
QY 115 VQYGGNNAALVNOTASDSSVMVRQV---GFGNNATAN 149
Db 301 TG-YG--NSGFYNAAVNNTGIFVTGVMSGGFFNFGTGN 335

RESULT 8
US-09-730-374-3
; Sequence 3, Application US/09730374
; Patent No. US20010031261A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US/09/730,374
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PCT/US99/12512
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,277
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A polypeptide encoded by an open reading frame of
; OTHER INFORMATION: SEQ ID NO:1
US-09-730-374-3

Query Match      11.3%; Score 87.5; DB 9; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 32; Indels 31; Gaps 2;

QY 27 GGGGNHNGGNSGPD-----STLSIYQGSANAAL 57
Db 126 GGGSGGGGGGGSDIELTQSPSSFSVSLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALQSDARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
Db 186 LLISGATSLETGVPFRFSGSGSKDYTLSTLSLTQTEDVA 224

RESULT 9
US-10-704-206-3
; Sequence 3, Application US/10704206
; Publication No. US20040141982A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES TO CD38 TO TREAT MULTIPLE
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2003-11-07
; OTHER INFORMATION: SEQ ID NO:1
US-10-704-206-3

Query Match      11.3%; Score 87.5; DB 9; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 32; Indels 31; Gaps 2;

QY 27 GGGGNHNGGNSGPD-----STLSIYQGSANAAL 57
Db 126 GGGSGGGGGGGSDIELTQSPSSFSVSLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALQSDARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
Db 186 LLISGATSLETGVPFRFSGSGSKDYTLSTLSLTQTEDVA 224

RESULT 10
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match      11.3%; Score 87.5; DB 15; Length 445;
Best Local Similarity 26.3%; Pred. No. 3.3;
Matches 45; Conservative 18; Mismatches 59; Indels 49; Gaps 8;

QY 7 AAFAA-----IVVSGSALAGVVPQWGGG-----NHNGG-----GNSGPDSTLSIYQY 50
Db 19 AAFADSNVTYVINGTNDQQANITQSGNGSVGVAFNGNSGFLQENGTLISGA-NLLTVKQS 77
QY 51 GSANAA-----LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQW 106
Db 78 GNSNSVGRDIOGKQSGAGNS-AAIFQEGTSDVLEQTGT-----SNGAVPSGW 125
QY 107 NAKN-----SDITVGYGGNNAALVNOTASDSSVMVRQVG 141
Db 126 NWTNDPGVFNKITQDSSNSGSKSVIODGKNVFSIKQGTGNTGNTSVNQG 176
```

```
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: GI1781260
US-09-820-843A-21

Query Match      11.5%; Score 89; DB 10; Length 354;
Best Local Similarity 24.1%; Pred. No. 1.7;
Matches 38; Conservative 17; Mismatches 63; Indels 40; Gaps 7;

QY 10 AAVVSGSALAG-VVPQWGGGNNHGG-----NSSGPDSTLSIYQGSAN 54
Db 200 AGDVNSGVNAGDVNTGLNSGNINTGFGNPGTLTGFSSAMTQAGFNS--GFENACTGN 257
QY 55 AALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDIT 114
Db 258 SGFGHNDPAGSGNGIQNSGFGNS-----GYVNTSTTSMFGNSGVLN 300
QY 115 VQYGGNNAALVNOTASDSSVMVRQV---GFGNNATAN 149
Db 301 TG-YG--NSGFYNAAVNNTGIFVTGVMSGGFFNFGTGN 335

RESULT 8
US-09-730-374-3
; Sequence 3, Application US/09730374
; Patent No. US20010031261A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES
; FILE REFERENCE: 150.188US2
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US/09/730,374
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PCT/US99/12512
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,277
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A polypeptide encoded by an open reading frame of
; OTHER INFORMATION: SEQ ID NO:1
US-10-704-206-3

Query Match      11.3%; Score 87.5; DB 16; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 32; Indels 31; Gaps 2;

QY 27 GGGGNHNGGNSGPD-----STLSIYQGSANAAL 57
Db 126 GGGSGGGGGGGSDIELTQSPSSFSVSLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALQSDARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
Db 186 LLISGATSLETGVPFRFSGSGSKDYTLSTLSLTQTEDVA 224

RESULT 10
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match      11.3%; Score 87.5; DB 15; Length 445;
Best Local Similarity 26.3%; Pred. No. 3.3;
Matches 45; Conservative 18; Mismatches 59; Indels 49; Gaps 8;

QY 7 AAFAA-----IVVSGSALAGVVPQWGGG-----NHNGG-----GNSGPDSTLSIYQY 50
Db 19 AAFADSNVTYVINGTNDQQANITQSGNGSVGVAFNGNSGFLQENGTLISGA-NLLTVKQS 77
QY 51 GSANAA-----LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQW 106
Db 78 GNSNSVGRDIOGKQSGAGNS-AAIFQEGTSDVLEQTGT-----SNGAVPSGW 125
QY 107 NAKN-----SDITVGYGGNNAALVNOTASDSSVMVRQVG 141
Db 126 NWTNDPGVFNKITQDSSNSGSKSVIODGKNVFSIKQGTGNTGNTSVNQG 176
```

```
RESULT 11
US-10-408-765A-998
; Sequence 998, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 998
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-998

Query Match      11.3%; Score 87.5; DB 16; Length 1448;
Best Local Similarity 28.6%; Pred. No. 14;
Matches 36; Conservative 16; Mismatches 39; Indels 35; Gaps 6;

QY 10 AAIVSGSALAGVVPQWGGG-GNHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSET 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 STLVDNGTAAGKPPSSGSGWGDHPA-----EPPVAFGRAGAPVAASALCKPEASKS-- 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 69 TITQSGYGGADYDQLVTRVTHEMAHAFRNATIDQNAKNSDITVQVGGNNAALVQ 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 633 --MOEGWGGSD-----EM-----NLSTQWEDEGDV-----WNNASQES 667

QY 129 TASDSS 134
   : : : : :
Db 668 TSSCSS 673

RESULT 12
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Ovendale, Pamela
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; TITLE OF INVENTION: of Tuberculosis
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mTTC#3-His
; US-09-793-306-146

Query Match      11.1%; Score 86; DB 12; Length 974;
Best Local Similarity 26.0%; Pred. No. 12;
Matches 39; Conservative 17; Mismatches 52; Indels 42; Gaps 7;

QY 15 SGSALAGVVPQWGGGNGGNGG-GNSSGPDSTLSIYQY-----SANA 55
```

```
Query Match      11.2%; Score 87; DB 9; Length 597;
Best Local Similarity 27.0%; Pred. No. 5.3;
Matches 34; Conservative 15; Mismatches 47; Indels 30; Gaps 6;

QY 26 WGGGNHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLV 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 FGNSGNNNIGFNSG-NNNVGFFNSGNNNFQGNAGD-----INTGFNGAGDTNT-- 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 86 TRVTHEMAHAFRNATIDQ--WNAKNSDITVQVGGNNAALVNOTASDSSVMVRQVGF 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 -----CFGNAGFFNMGIGNAGNEDMGVGGSGSFNVGVGN--AGNQS-----VGFG 449
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 144 NNATAN 149
   : : : : :
Db 450 NAGTLN 455
```

```
RESULT 13
US-10-282-122A-44999
; Sequence 44999, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44999
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-10-282-122A-44999
```

```
Query Match      11.1%; Score 86; DB 12; Length 974;
Best Local Similarity 26.0%; Pred. No. 12;
Matches 39; Conservative 17; Mismatches 52; Indels 42; Gaps 7;

QY 15 SGSALAGVVPQWGGGNGGNGG-GNSSGPDSTLSIYQY-----SANA 55
```

Db 299 AGNGA-----SGNGHNYGNGGDDVITAPITGVNLISGNSFTLIGNSSSSVNT 352
QY 56 ALALQSDARKSETTI-----TSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNA 108
Db 353 APTTNTVNDNDITDNGNSGCTGSGGSG-DLLNGAASNGER----NYGNGNG 407
QY 109 KNSDIT-----VGQVGNNAALVNOTASDS 133
Db 408 DDVITAPITGVNFSGNSFSLIGNSSSS 437

RESULT 14

US-10-282-122A-64364
; Sequence 64364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELTRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/205,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 64364
; LENGTH: 2204
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-64364

Query Match 11.0%; Score 85.5; DB 12; Length 2204;
Best Local Similarity 25.0%; Pred. No. 38;
Matches 34; Conservative 15; Mismatches 56; Indels 31; Gaps 4;
QY 30 GNHNGGNSGPDST-----LSIYQGSANAALQSDARKSETTITQSGYNGADYDQLV 85
Db 1886 GGVNIGNNGIANTGTFDGLNLGNSYNGLANLGD-----DNLGFGNAGSYNIGF 1937
QY 86 TRVTHEMAHAFRNATIDOWNAKNSDITVG-----CYGNNAAALVNOTASDS 133
Db 1938 ANFGSDNLGPANTGSYNIGFANTGNNIGVLTGNGQIGIGLSNGSNNIGLFGNSGSGN- 1996

QY 134 SVMVRQVGFNNATAN 149
Db 1997 -----IGFFNSGTGN 2006

RESULT 15

US-09-880-748-905
; Sequence 905, Application US/09880748
; Publication No. US20030059937A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PFS23
; CURRENT APPLICATION NUMBER: US/09/880,748
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 905
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-905

Query Match 11.0%; Score 85; DB 10; Length 254;
Best Local Similarity 23.0%; Pred. No. 2.9;
Matches 29; Conservative 11; Mismatches 42; Indels 44; Gaps 3;

QY 13 VVSGSALAGVWPONG-----GGNHNGGNSGPDSTLSI----- 47
Db 104 ILTYMGSAFDQWGRGTMTVSSGGGGGGGGGSALEIVNTQSPGTLISLSPGERA 163
QY 48 -----YQGSANAALQSDARKSETTITQ--SGYNGADYDQLVTRV 88
Db 164 TLSCRASQSIIRSNLYAWYQKSGQAPRLLIYDVSSRATGIPDRFSGSGSGTDFLTISRL 223
QY 89 VTHEMA 94
Db 224 EPEDFA 229

Search completed: August 2, 2004, 15:36:13
Job time : 36.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.909 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAAPRAIVVSGSALA.....DSSVMVRQVFGNNATANQV 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA Main:*

1: /cgn2_6/ptodata/2/paa/PCRUS_COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US09_COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A_COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B_COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098_COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A_COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B_COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100_COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101_COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102_COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103_COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104_COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106_COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107_COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query Match | Score | Length | DB ID | Description |
|------------|-------------|-------|--------|-------|-------------|
|------------|-------------|-------|--------|-------|-------------|

| | | | | | | |
|----|-------|-------|------|----|----------------------|--------------------|
| 1 | 775 | 100.0 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 2 | 712 | 91.9 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 3 | 683 | 88.1 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 4 | 678 | 87.5 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 5 | 662 | 85.4 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 6 | 617 | 79.6 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 7 | 613 | 79.1 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
| 8 | 605 | 78.1 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 9 | 603 | 77.8 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 10 | 601 | 77.5 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 11 | 600 | 77.4 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 12 | 596 | 76.9 | 131 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 13 | 568 | 73.3 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 509 | 65.7 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 506 | 65.3 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 506 | 65.3 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 504 | 65.0 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 504 | 65.0 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 498 | 64.3 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 464 | 59.9 | 109 | 19 | US-09-543-407-34 | Sequence 34, Appl |
| 21 | 462 | 59.6 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 22 | 462 | 59.6 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 23 | 462 | 59.6 | 158 | 30 | US-10-417-886-5834 | Sequence 5834, Ap |
| 24 | 428 | 55.2 | 131 | 13 | US-08-978-878-2 | Sequence 2, Appl |
| 25 | 428 | 55.2 | 131 | 21 | US-09-741-873B-2 | Sequence 35, Appl |
| 26 | 324 | 41.8 | 109 | 19 | US-09-543-407-37 | Sequence 37, Appl |
| 27 | 257 | 33.2 | 68 | 19 | US-09-543-407-39 | Sequence 39, Appl |
| 28 | 181 | 23.4 | 48 | 19 | US-09-543-407-32 | Sequence 32, Appl |
| 29 | 160.5 | 20.7 | 70 | 19 | US-09-543-407-33 | Sequence 5833, Ap |
| 30 | 102.5 | 13.2 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 31 | 102.5 | 13.2 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 102.5 | 13.2 | 186 | 30 | US-10-417-886-5833 | Sequence 18734, A |
| 33 | 95 | 12.3 | 400 | 21 | US-09-733-089-18734 | Sequence 18734, A |
| 34 | 95 | 12.3 | 400 | 23 | US-09-816-660-18734 | Sequence 28424, A |
| 35 | 95 | 12.3 | 400 | 27 | US-10-158-881-28424 | Sequence 186417, A |
| 36 | 95 | 12.3 | 400 | 30 | US-10-437-963-186417 | Sequence 17364, A |
| 37 | 95 | 12.3 | 400 | 30 | US-10-438-246-17264 | Sequence 18735, A |
| 38 | 95 | 12.3 | 1060 | 21 | US-09-733-089-18735 | Sequence 18735, A |
| 39 | 95 | 12.3 | 1060 | 23 | US-09-816-660-18735 | Sequence 6, Appl |
| 40 | 94.5 | 12.2 | 151 | 19 | US-09-543-407-6 | Sequence 41783, A |
| 41 | 92.5 | 11.9 | 644 | 22 | US-09-791-537-108562 | Sequence 108562, A |
| 42 | 92 | 11.9 | 598 | 22 | US-09-791-537-108562 | Sequence 50616, A |
| 43 | 92 | 11.9 | 1074 | 28 | US-10-282-122A-50616 | Sequence 23883, A |
| 44 | 91.5 | 11.8 | 478 | 22 | US-09-791-537-23883 | Sequence 29231, A |
| 45 | 91.5 | 11.8 | 602 | 22 | US-09-791-537-29231 | |

ALIGNMENTS

RESULT 1
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Ray, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-28

```
Query Match      100.0%; Score 775; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.2e-74;
Matches 151; Conservative 0; Mismatches 0; Indels 0
```

| | | | | | | | | |
|-----|---------------|----------------|----------------|----------------|----------------|------------|---------|---------|
| 1 | MKLIKVAAPFAAI | VVGSALAGVVP | QWGGGNGHGGNS | GGPDS | TLSTI | YQVGS | NAALAIQ | 60 |
| | 1 | MKLIKVAAPFAAI | VVGSALAGVVP | QWGGGNGHGGNS | GGPDS | TLSTI | YQVGS | NAALAIQ |
| 61 | SDARKSETTIT | IQSGYNGADY | QDLVTRV | VVTHEMAHAFENNA | TIDQWAKNSD | ITVQYGG | 120 | |
| | 61 | SDARKSETTIT | IQSGYNGADY | QDLVTRV | VVTHEMAHAFENNA | TIDQWAKNSD | ITVQYGG | |
| 121 | NNAALVNQTAS | DSVMVROVGFENNA | TANQY | 151 | | | | |
| | 121 | NNAALVNQTAS | DSVMVROVGFENNA | TANQY | 151 | | | |

RESULT 2

US-09-543-407-22

```

: Sequence 22, Application US/09543407
:
: GENERAL INFORMATION:
:
: APPLICANT: White, Aaron P.
: APPLICANT: Doran, James L.
: APPLICANT: Collinson, S. Karen
: APPLICANT: Kay, William W.
:
: TITLE OF INVENTION: BACTERIAL FIMBRIAL SVSTM FOR
: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
:
: TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
: FILE REFERENCE: 920043.406
: CURRENT APPLICATION NUMBER: US/09/543,407
: CURRENT FILING DATE: 2000-04-05
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 22
: LENGTH: 151
: TYPE: PRT
:
: FEATURE:
:
: OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
: OTHER INFORMATION: sequence containing the replacement fragment
: OTHER INFORMATION: encoding P73 from GPe3 of Leishmania major.
: US-09-543,407-22

```

```

Query Match      91.9%; Score 712; DB 19; Length 151;
Best Local Similarity 91.1%; Prid.No. 7.2e-68;
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAFAAAIVVSGSALAGVWPQGGGNGHNGGNSGDPSTLSIYYGSANAALALQ 60
Db 1 MKLLKVAFAAAIVVSGSALAGVWPQGGGNGHNGGNSGDPSTLSIYYGSANAALALQ 60

QY 61 SPARKSETTITQSGYNGAD-----YDGLVTRVVVTHEMAHAFNNATTIQWNAKNSDI 113
Db 61 SPARKSETTITQSGYNGADVGGADNYDGLVTRVVVTHEMAH-----DQWNAKNSDI 113

```

Qy

114 TVGQYGGNNAALVNQTASDSSVMVRQVVGFGNNATANQY 151
|||||
|||

Dβ

114 TVGQYGGNNAALVNQTASDSSVMVRQVVGFGNNATANQY 151
|||||
|||

RESULT 3

US-09-543-407-5

```

; Sequence 5, Application US/09543407
;
; GENERAL INFORMATION:
;
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
;
; TITLE OF INVENTION: BACTERIAL FILMIRAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
;

```

; CURRENT FILING DATE: 2000-04-05

```

; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151

```

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------|
| Query Match | 88.1%; | Score 683; | DB 19; | Length 151; |
| Best Local Similarity | 90.1%; | Pred. No. 9.5e-85; | | |
| Matches 136: | Conservative | 1; | Mismatches 14; | Indels 0; |
| | Capable | 0; | | |

[illegible]

RESULT 4

US-08-233-642A-57

```

1 Sequence 57, Application US/08233642A
2 GENERAL INFORMATION:
3 APPLICANT: Kay, William W.
4 APPLICANT: Collinson, S. Karen
5 APPLICANT: Clouthier, Sharon C.
6 APPLICANT: Doran, James L.
7 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
8 BASED VACCINES
9 NUMBER OF SEQUENCES: 58
10 NUMBER OF SEQUENCES: -
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Seed and Berry
13 STREET: 6300 Columbia Center, 701 Fifth Avenue
14 CITY: Seattle
15 STATE: Washington
16 COUNTRY: U.S.A.
17 ZIP: 98104-7092
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patent in Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/233,642A
25 FILING DATE: 26-APR-1994
26 CLASSIFICATION: 424
27 ATTORNEY/AGENT INFORMATION:
28 NAME: King, Joshua
29 REGISTRATION NUMBER: 35,570
30 REFERENCE/DOCKET NUMBER: 920043.403C3
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (206) 622-4900
33 TELEFAX: (206) 682-6031
34 TELEX: 3723836 SEEDANBERRY
35 INFORMATION FOR SEQ ID NO: 57:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 151 amino acids
38 TYPE: amino acid
39 TOPOLOGY: linear
40 MOLECULE TYPE: protein
41 US-08-233-642A-57

```

Query Match 87.5%; Score 678; DB 6; Length 151;
Best Local Similarity 89.4%; Pred. No. 3.3e-64;
Matches 135; Conservative 1; Mismatches 15; Indels

FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407

```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVGYGG 120
Dd 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
Dd 121 NNPALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 5
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20
Query Match 85.4%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 1.7e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQ 105
Dd 61 SDARK-----YDQLVTRVVTHEMAFRNNATIDQ 105
QY 106 WNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151
Dd 106 WNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151

RESULT 6
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26
Query Match 79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAFR 98
Dd 61 SDARKSETTITQSGYNGADNNTIELTQNGFRNNATYDQLVTRVVTHEMAFR 118
QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151
Dd 119 -----GGNNAALVNQTSASSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30
Query Match 79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAFR 98
Dd 61 SDARKSETTITQSGYNGADNNTIELTQNGFRNNATYDQLVTRVVTHEMAFR 118
QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151
Dd 119 -----GGNNAALVNQTSASSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-26
Query Match 79.6%; Score 617; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.2e-57;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 57
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFR-----F 97
Dd 58 -----YDQLVTRVVTHEMAFRAGYNGADVQGGADNSTIETQNGF 97
QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151
Dd 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151

RESULT 7
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-30
Query Match 79.1%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 3.2e-57;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
Dd 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAFR 98
Dd 61 SDARKSETTITQSGYNGADNNTIELTQNGFRNNATYDQLVTRVVTHEMAFR 118
QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVMVRQVGFNNATANQY 151
Dd 119 -----GGNNAALVNQTSASSVMVRQVGFNNATANQY 151

RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 78.1%; Score 605; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.3e-56;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 9
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match 77.8%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.8e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120

Query Match 77.4%; Score 600; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;
```

```
QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 10
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-18

Query Match 77.5%; Score 601; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 6.3e-56;
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
DB 61 SDARKSETTITQSGYNGADVDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 11
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-24

Query Match 77.4%; Score 600; DB 19; Length 151;
```

Best Local Similarity 81.5%; Pred. No. 8.1e-56;
Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120

RESULT 12

US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 76.9%; Score 596; DB 19; Length 131;
Best Local Similarity 88.5%; Pred. No. 1.8e-55;
Matches 116; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 YDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGNNALVNQTASDSSVMVRQV 140
Db 61 VGQADNSTIETQNGFRNATIDOWNAKNSDITVGYGNNALVNQTASDSSVMVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 13

US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16

Query Match 73.3%; Score 568; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.2e-52;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 14

US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 65.7%; Score 509; DB 19; Length 151;
Best Local Similarity 68.2%; Pred. No. 5e-46;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGADYDQGGSDSDSIDLTQGFNGSATLDOWNKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATANQY 151

RESULT 15

US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06

Tue Aug 3 10:54:44 2004

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      65.3%; Score 506; DB 13; Length 151;
Best Local Similarity 67.5%; Pred. No. 1e-45;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALAQ 60
Db 1 MKLLKVAAPAAIVFSGSAGVVPQYGGGNGHNGGNSGPNSEINIIYQGGNSALAQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFENNATIDQWAKNSDITVQYGG 120
Db 61 TDARNSDLTITQHGCGNGADYQGGSDSSIDLQKFCNSATLDQWNGKXSEMTVKQFGG 120
QY 121 MNAALVNOTASDSSVMVQVGFNNATANQY 151
Db 121 GNGAAVDQATASNSVNVTVQVGFNNATAHQY 151

```

Search completed: August 2, 2004, 15:26:46
Job time : 167.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pcp:**
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pcp:**
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pcp:**
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pcp:**
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pcp:**
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pcp:**
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pcp:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 506 | 65.3 | 151 | 5 | US-09-741-873C-4 |
| 2 | 428 | 55.2 | 131 | 5 | US-09-741-873C-2 |
| 3 | 89.5 | 11.5 | 388 | 5 | US-09-248-796A-17306 |
| 4 | 87.5 | 11.3 | 619 | 1 | PCT-US04-10229-38 |
| 5 | 87.5 | 11.3 | 1905 | 1 | PCT-US04-09388-9 |
| 6 | 84.5 | 10.9 | 234 | 6 | US-10-787-701-45603 |
| 7 | 84.5 | 10.9 | 511 | 6 | US-10-425-115-320950 |
| 8 | 82 | 10.6 | 276 | 6 | US-10-425-115-339993 |
| 9 | 81 | 10.5 | 147 | 6 | US-10-425-115-193207 |
| 10 | 81 | 10.5 | 224 | 6 | US-10-805-394-4894 |
| 11 | 81 | 10.5 | 457 | 5 | US-09-248-796A-26659 |
| 12 | 81 | 10.5 | 568 | 1 | PCT-US04-21492-458 |
| 13 | 80 | 10.3 | 1627 | 6 | US-10-170-205E-16559 |
| 14 | 79 | 10.2 | 295 | 6 | US-10-425-115-312468 |
| 15 | 79 | 10.2 | 1010 | 1 | PCT-US04-21492-402 |
| 16 | 78 | 10.1 | 892 | 5 | US-09-952-267B-5 |
| 17 | 78 | 10.1 | 892 | 6 | US-10-872-769-5 |
| 18 | 78 | 10.1 | 892 | 6 | US-10-872-769-5 |
| 19 | 77.5 | 10.0 | 246 | 6 | US-10-854-439-511 |
| 20 | 77.5 | 10.0 | 415 | 5 | US-10-490-953-13 |
| 21 | 77.5 | 10.0 | 520 | 6 | US-10-479-638-21 |
| 22 | 77 | 9.9 | 1160 | 6 | US-10-603-150-2 |
| 23 | 76.5 | 9.9 | 132 | 6 | US-10-425-115-351875 |
| 24 | 76.5 | 9.9 | 248 | 6 | US-10-170-205E-3429 |
| 25 | 76.5 | 9.9 | 283 | 6 | US-10-854-439-1222 |
| 26 | 76.5 | 9.9 | 685 | 5 | US-09-248-796A-15216 |

| | | | | | | |
|----|------|-----|------|---|----------------------|-----------------------|
| 27 | 76.5 | 9.9 | 873 | 5 | US-09-952-267B-13 | Sequence 13, Appl |
| 28 | 76.5 | 9.9 | 873 | 6 | US-10-872-768-13 | Sequence 13, Appl |
| 29 | 76.5 | 9.9 | 873 | 6 | US-10-872-769-13 | Sequence 13, Appl |
| 30 | 76.5 | 9.9 | 1058 | 6 | US-10-796-907-72 | Sequence 72, Appl |
| 31 | 76 | 9.8 | 127 | 6 | US-10-425-115-274032 | Sequence 274032, Appl |
| 32 | 76 | 9.8 | 201 | 6 | US-10-425-115-309662 | Sequence 309662, Appl |
| 33 | 76 | 9.8 | 279 | 6 | US-10-425-115-343835 | Sequence 343835, Appl |
| 34 | 76 | 9.8 | 598 | 6 | US-10-425-115-221471 | Sequence 221471, Appl |
| 35 | 76 | 9.8 | 688 | 7 | US-60-581-351-11388 | Sequence 11388, A |
| 36 | 76 | 9.8 | 753 | 7 | US-60-581-351-12224 | Sequence 12224, A |
| 37 | 76 | 9.8 | 822 | 6 | US-10-425-115-287789 | Sequence 287789, Appl |
| 38 | 76 | 9.8 | 3132 | 7 | US-60-556-841-1245 | Sequence 1245, Appl |
| 39 | 75.5 | 9.7 | 400 | 6 | US-10-490-953-13 | Sequence 13, Appl |
| 40 | 75.5 | 9.7 | 400 | 6 | US-10-490-953-14 | Sequence 14, Appl |
| 41 | 75.5 | 9.7 | 400 | 6 | US-10-490-953-20 | Sequence 20, Appl |
| 42 | 75.5 | 9.7 | 400 | 6 | US-10-723-981-14 | Sequence 14, Appl |
| 43 | 75.5 | 9.7 | 400 | 6 | US-10-723-981-15 | Sequence 15, Appl |
| 44 | 75.5 | 9.7 | 424 | 6 | US-10-045-674A-591 | Sequence 591, Appl |
| 45 | 75.5 | 9.7 | 529 | 6 | US-10-425-115-340009 | Sequence 340009, Appl |

ALIGNMENTS

RESULT 1

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

Query Match 65.3%; Score 506; DB 5; Length 151;
Best Local Similarity 67.5%; Pred. No. 6.1e-38;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQVGGNGGNGGNSGDPSTLSIYQYGSANAALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQVGGNGGNGGNSGDPSTLSIYQYGSANAALQ 60
Qy 61 SPARKSETTITQSGYNGGADYDQLVTRVVTHEMAFAFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 TARNSDLTITQGGNGGADVQGGDDSSIDYTORGFNGSATLDOWNKNSNTVQKQFG 120
Qy 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151
Db 121 GNGAAVDQTASNSSVNVTVQVFGNNATAHQY 151

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-08-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-741-873C-2

Query Match 55.2%; Score 428; DB 5; Length 131;
Best Local Similarity 64.1%; Pred. No. 5e-31;
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;

QY 21 GVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGAD 80
DB 1 GVVPQGGGNGHNGGNSGPNSELTQYGGGNSALALQTDARNSDITIQHGGNGAD 60
QY 81 YDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVQYGGNNAALVNQTSADSSVMVQV 140
DB 61 VQGGSDSSIDITQRFNGSALTDQWNGKNSMTVKQFGGNGAAGVDTASNSVNVTVQ 120
QY 141 GFGNNATANQY 151
DB 121 GFGNNATANQY 131

RESULT 3
US-09-248-796A-17306
; Sequence 17306, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17306
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-17306

Query Match 11.5%; Score 89.5; DB 5; Length 388;
Best Local Similarity 22.1%; Pred. No. 3.6;
Matches 33; Conservative 17; Mismatches 44; Indels 55; Gaps 4;

QY 30 GNHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADYDQLVTRV 89
DB 132 GNQVGSNSYSPDT-----YGSAITGLGQEKTPAVTGIHSGIG-AAAYPELT---- 180
QY 90 THEMAHAFRNATIDQWNAKNSDITVQYGGNNAALVNQTS----- 131
DB 181 -----NAGNTGLAKGTAPASTSATYGESPSADYSKSGATGVVPATYLN 223
QY 132 -----DSSVMVQVQFGNNATANQY 151
DB 224 TSCAPTGSLNTAGVGGAGFGDNTSSY 252

RESULT 4
PCT-US04-10229-38
; Sequence 38, Application PC/TUS0410229
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Kerovuo, Janne
; APPLICANT: Solbak, Arne
; APPLICANT: Gray, Kevin
; APPLICANT: McCann, Ryan
; APPLICANT: Purohit, Shalaka
; APPLICANT: Gerendash, Joel
; APPLICANT: Janssen, Giselle
; APPLICANT: Dahod, Samun
; TITLE OF INVENTION: PECTATE LYASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462009640
; CURRENT APPLICATION NUMBER: PCT/US04/10229
; CURRENT FILING DATE: 2003-04-03
; PRIOR APPLICATION NUMBER: 60/460,842
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: 60/484,798
; PRIOR FILING DATE: 2003-07-03
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: SIGNAL
; LOCATION: (1)...(35)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (36)...(387)
; OTHER INFORMATION: Catalytic domain
; PCT-US04-10229-38

Query Match 11.3%; Score 87.5; DB 1; Length 619;
Best Local Similarity 25.9%; Pred. No. 9.4; Indels 45; Gaps 4;
Matches 36; Conservative 10; Mismatches 43;

QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQ 61
DB 18 LTAIAAVALMAGTSFAAAT---GGFSTTDGGNVSGARSFTASTYQQINTIIANAKLD 73
QY 62 DARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVQYGCN 121
DB 74 DAGKVT-----GGAYPLIIT-----YTGN 93
QY 122 NAALVNQTSADSSV 135
DB 94 EDSLINQMIKDHV 107

RESULT 5
PCT-US04-09388-9
; Sequence 9, Application PC/TUS0409388
```



```
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; MARQUIS, Joseph P.;
; APPLICANT: TRAN, Uyen K.; YANG, Yonghong G.;
; APPLICANT: TANG, Y. Tom; CHAWLA, Narinder K.;
; APPLICANT: MURAGE, Jaji; WANG, Jonathan T.;
; APPLICANT: ELLIOTT, Vicki S.; CHIEN, David;
; APPLICANT: YUE, Henry; AZIMZAI, Yalda;
; APPLICANT: ISON, Craig H.; KHARE, Reena;
; APPLICANT: JIN, Pei; RAMKUNAR, Jayalaxmi;
; APPLICANT: FAVERO, Kristin D.; RICHARDSON, Thomas W.;
; APPLICANT: HAFALIA, April J.A.; BAUGHN, Mariah R.;
; APPLICANT: BECHA, Shanva D.; WILSON, Amy D.
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PF-1531 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/09388
; CURRENT FILING DATE: 2004-04-01
; PRIOR APPLICATION NUMBER: US 60/457,403
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/465,568
; PRIOR FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: US 60/476,135
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 60/476,583
; PRIOR FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 1905
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7525307CD1
PCT-US04-09388-9

Query Match 11.3%; Score 87.5; DB 1; Length 1905;
Best Local Similarity 28.6%; Pred. No. 35;
Matches 36; Conservative 16; Mismatches 39; Indels 35; Gaps 6;

Qy 10 AAIIVGSGALAGVVPQGGG-GNHNGGNSGPDSTLSIYQGSANAALQSDARKSET 68
Db 1031 STLVDNGTAAGKPPSSGSGWDHPA-----BPPVAFGRAGAPVAASALCKPASKS-- 1081
Qy 69 TITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
Db 1082 --MQEGWSSGD-----EM-----NLSTSQWEDEEGDV-----WNAASQES 1116

Qy 129 TASDSS 134
Db 1117 TSSCSS 1122

RESULT 6
US-10-767-701-45603
; Sequence 45603, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45603
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(234)
; OTHER INFORMATION: unsure at all Xaa locations

; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; MARQUIS, Joseph P.;
; APPLICANT: TRAN, Uyen K.; YANG, Yonghong G.;
; APPLICANT: TANG, Y. Tom; CHAWLA, Narinder K.;
; APPLICANT: MURAGE, Jaji; WANG, Jonathan T.;
; APPLICANT: ELLIOTT, Vicki S.; CHIEN, David;
; APPLICANT: YUE, Henry; AZIMZAI, Yalda;
; APPLICANT: ISON, Craig H.; KHARE, Reena;
; APPLICANT: JIN, Pei; RAMKUNAR, Jayalaxmi;
; APPLICANT: FAVERO, Kristin D.; RICHARDSON, Thomas W.;
; APPLICANT: HAFALIA, April J.A.; BAUGHN, Mariah R.;
; APPLICANT: BECHA, Shanva D.; WILSON, Amy D.
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PF-1531 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/09388
; CURRENT FILING DATE: 2004-04-01
; PRIOR APPLICATION NUMBER: US 60/457,403
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/465,568
; PRIOR FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: US 60/476,135
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 60/476,583
; PRIOR FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 1905
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7525307CD1
PCT-US04-09388-9

Query Match 11.3%; Score 87.5; DB 1; Length 1905;
Best Local Similarity 28.6%; Pred. No. 35;
Matches 36; Conservative 16; Mismatches 39; Indels 35; Gaps 6;

Qy 10 AAIIVGSGALAGVVPQGGG-GNHNGGNSGPDSTLSIYQGSANAALQSDARKSET 68
Db 1031 STLVDNGTAAGKPPSSGSGWDHPA-----BPPVAFGRAGAPVAASALCKPASKS-- 1081
Qy 69 TITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVQYGGNNAALVNQ 128
Db 1082 --MQEGWSSGD-----EM-----NLSTSQWEDEEGDV-----WNAASQES 1116

Qy 129 TASDSS 134
Db 1117 TSSCSS 1122

RESULT 6
US-10-767-701-45603
; Sequence 45603, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 45603
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(234)
; OTHER INFORMATION: unsure at all Xaa locations
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82834_1.pep
US-10-767-701-45603

Query Match 10.9%; Score 84.5; DB 6; Length 234;
Best Local Similarity 24.2%; Pred. No. 5.5; Indels 65; Gaps 6;
Matches 39; Conservative 11; Mismatches 46;

Qy 1 MKLLKVAFAAIVV--SGSALAGVVPQW---GGGGNHNGGNN-----SSGPDSTL 45
Db 1 MATTKLAALCFIVLLGIGGANAARVARYVSAGGGGGGGGGRWRGGASRWGSGSGC 60
Qy 46 SIY-----OYGSANAALALQSDARKSETTITQSGYNGADYD 82
Db 61 GXYGEAGSGXAYAQGGGGGGGGGGGSG-----SGYSGSGYG 102
Qy 83 QLTVRVVTHEMAHAFRNATIDQWNAKNSDITVQYGGNNA 123
Db 103 Q-----AGSGSNGGAYAQGGAQGGGGGGGGYGGSGS 134

RESULT 7
US-10-425-115-320950
; Sequence 320950, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 320950
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(511)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: MRT4577_55769C.1.pep
US-10-425-115-320950

Query Match 10.9%; Score 84.5; DB 6; Length 511;
Best Local Similarity 23.3%; Pred. No. 14;
Matches 44; Conservative 22; Mismatches 64; Indels 59; Gaps 8;

Qy 1 MKLLKVAFAAIVWS-----GSALAGVVPQWGG-----GNHNGGNSGPD 42
Db 133 LRLLEERAQEAIALASCRDVLRSAGFFEDAWAKVIPGSDEGVAVVAANYALQRLGGDPN 192
Qy 43 STLSTIYQGSANAALALQSDA-----RKSETTITQ-----SGYNGADYDQLVTR 87
Db 193 KTVGIIELGCSAQLTFVSDEVLPPKLSYNYTFGETTYLTNTSFLNFGNAQAQDSF--- 249
Qy 88 VVTHEVAH---AFRNATID---QWNAKNSDITVQYGGNNAALVNQASDSSVWVQV 140
Db 250 ---HEMLRGRGFKNGTLADPCAPRGYSRNEEMLRMSGASRSTLENQ----- 294
Qy 141 GFQNNATAN 149
Db 295 -YVNNGTGN 302

RESULT 8
US-10-425-115-339993
; Sequence 339993, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 339993
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(276)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_73240C.1.pep
;
US-10-425-115-339993

Query Match 10.6%; Score 82; DE 6; Length 276;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 29; Conservative 17; Mismatches 33; Indels 28; Gaps 5;

Qy 12 IVVSGAAGVVPWGCGGNGHNGSGSPDSTLSIYCYGSANAALQSDARKSETTIT 71
Db 10 LVLVGVSGCGCAQGGGG-----GGGLRG--SFPKGFVFGIA-----SSAYQIEGAVK 56

Qy 72 QSGYGNMGADYDGLVTRVVTVMHAFRNATIDOWNAKNSDITVGQY 118
Db 57 EDGRG-----KTIWDKFAITFGKVADL-----SNADVADVOY 88

```

[illegible]

RESULT 10

US-10-805-394-4894

; Sequence 4894, Application US/10805394

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/10/805,394

; CURRENT FILING DATE: 2004-03-22

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988

; PRIOR FILING DATE: 2000-08-03

; NUMBER OF SEQ ID NOS: 7059

; SOFTWARE: PatentIn ver. 3.0

; SEQ ID NO 4894

; LENGTH: 224

; TYPE: PRT

; ORGANISM: Corynebacterium glutamicum

US-10-805-394-4894

Query Match: 10.5%; Score 81; DB 6; Length 224;

Best Local Similarity 24.6%; Pred. No. 11;

Matches 35; Conservative 19; Mismatches 54; Indels 34; Gaps 4;

QY 16 GSAAGVWPQWGGGNNHNGGNS-----GPDSTLSIYQYGSANALALQSDARKSETTIT 71

Db 45 GTAAGVLTGKGATVEHIGSADSIATINGDVVILAVPYPAVESITASHKDALAGKTVID 104

QY 72 QSGYNGAGDYDOLV-----TRVTHEMAHAFRNN--ATI----- 103

Db 105 INPLNFETDSLVVPGSSATAETAOQLPTSRVLKAFNTNPAAILATKVGVDITTVLV 164

QY 104 --DOWNAKNSDITVGQYGGNNA 123

Db 165 AGDDEDAKVALITDVNAGGLDA 186

RESULT 11

US-09-248-796A-26659

; Sequence 26659, Application US/09248796A

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 26659

; LENGTH: 457

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-26659

Query Match: 10.5%; Score 81; DB 5; Length 457;

Best Local Similarity 25.0%; Pred. No. 25;

Matches 40; Conservative 19; Mismatches 69; Indels 32; Gaps 6;

QY 15 SGSALAGVYVPMWGGGNGHGGNS-----GPDSTLSIYQYGSANAALALQSDARKSET 68
Db 289 SSEFLASLINGLGGGGGNGSGNTSNYKHNSTTSTTSKYFNSSTATKLSSSKSIYVNS 348
QY 69 TITQSGY-----GNGADYDOL-----VTRVVTHEMAHAFRNATIDQWNAKSD 112
Db 349 TTSRSSLSVSSSTDDGGGANLFGLLNSVAASVSRTLAAESTLSTGTTTSDSANGTKD 408
QY 113 ITVQYGGNNAALVNQTASDSSVMVRQVGFQNNAT-ANQY 151
Db 409 YS--SYSGTITSFSTTCSLS-----GDGNKLGNGKY 439

RESULT 12

PCT-US04-21492-458
; Sequence 458, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 458
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from environmental sample
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(39)
; NAME/KEY: DOMAIN
; LOCATION: (46)...(307)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (470)...(567)
; OTHER INFORMATION: Cellulose binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (385)...(413)
; OTHER INFORMATION: Cellulose or protein binding domain
PCT-US04-21492-458

Query Match 10.5%; Score 81; DB 1; Length 568;
Best Local Similarity 20.7%; Pred. No. 32;
Matches 35; Conservative 22; Mismatches 52; Indels 60; Gaps 7;
QY 26 WGGGNGHN-----GGNNSGPDSTLSIYQYGSANAALALQSDARKSET 68
Db 403 WGNEGNCVAVSTCSSQPAPYGVVGGSSSGMSSSVAPSSSSSTVSLPASSSSSS 462
QY 69 TITQSGYNGADYDOLVTRVVTHEMAHAF-----RNNATIDQWNAK----- 109
Db 463 SSTASGDGN-CEY-----IVSNEWNTGFTCAIRITNEGSAINGNVNSYSYDGTSVTS 515
QY 110 -----NSDITVQYGGNNAALVNQTASDSSVMVRQVGF-GNNATAN 149
Db 516 AWNATVSGNNPYASNLGNWNGTIQPGQTV-----EFGQGNKGSN 556

RESULT 13

US-10-170-205E-16659
; Sequence 16659, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CLO01381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 16659
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-16659

Query Match 10.3%; Score 80; DB 6; Length 1627;
Best Local Similarity 26.3%; Pred. No. 1.4e+02;
Matches 41; Conservative 16; Mismatches 51; Indels 48; Gaps 9;
QY 10 AAIIVSGSALAGVVPQ-----WGGGNGHN-----GGNNSGPDSTLSI 47
Db 621 AAAAXSGHAWSGAANQEDKSPTWGPPKPKSCHWGDGQSRNPASGAGGDWADSSSVLGH 680
QY 48 YOYGSANAALALQSDARKS-----ETTTT-QSGYNGADYDOLVTRVVTHEMAHAFRN-N 100
Db 681 LGDGKKNGS-GWDALNSRSGSGWNDDTRSGNSGWNS-----TNTKANFGTNWG 728
QY 101 ATI-----DQWNAKSDITVQYGGNNAALVNQTAS 131
Db 729 ETLKPGPQQNWASKPDNNVSNWGG--AASVKQTGT 762

RESULT 14

US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pap
US-10-425-115-312468

Query Match 10.2%; Score 79; DB 6; Length 295;
Best Local Similarity 25.4%; Pred. No. 23;
Matches 32; Conservative 13; Mismatches 43; Indels 38; Gaps 5;
QY 26 WGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDOLV 85
Db 138 YGGGGYSGGGYSSG-GYAANGYGVGSGNYSNAGGGYSGS---DGYNGA----- 186
QY 86 TRVVTHEMAHAFRNATIDQWNAKSDITVQYGGNNAALVNQTASDSSVMVRQVGFQNN 145
Db 187 -----ASGGYANNLS-SGYNNGRYNTIGSSDGTG-----GYN 219
QY 146 ATANQY 151
Db 220 SYPNPY 225

```
RESULT 15
PCT-US04-21492-402
; Sequence 402, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; CURRENT FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 402
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(30)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (393)...(428)
; OTHER INFORMATION: Carbohydrate binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (39)...(300)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (610)...(959)
; OTHER INFORMATION: Glycosyl hydrolases family 6
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (493)...(521)
; OTHER INFORMATION: Cellulose or protein binding domain
PCT-US04-21492-402

Query Match 10.2%; Score 79; DB 1; Length 1010;
Best Local Similarity 20.8%; Pred. No. 96;
Matches 32; Conservative 27; Mismatches 45; Indels 50; Gaps 7;

QY 12 IVVGSALAGVVPQWG--GGGNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSET 68
Db 316 LTASGLVKNIVSNWGTIGNGSSSSSSSSSSSS-----SSSSSSSSSSSSSSSS 367
QY 69 TITQSGYGN-----GADY-----DQLVTRVVTHEMAHAFRNATIDQW 106
Db 368 SSGSTGGGNCAGVNVYPNWTARDWSGGAYNHANAGDQW-----YQNSLYRANW 416
QY 107 ---NAKNSD---ITVQYGGNNAALVNOTASDSS 134
Db 417 YTNVPGSDASWTSGLGACGGNGSTSSSSSSSS 450
```

Search completed: August 2, 2004, 15:29:55
Job time : 17.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Query | Score | Match | Length | DB ID | Description |
|------------|-------|-------|-------|--------|--------|--------------------|
| 1 | 683 | 88.1 | 151 | 2 | JC6039 | fibrin protein ag |
| 2 | 683 | 88.1 | 151 | 2 | A10635 | major curlin chain |
| 3 | 509 | 65.7 | 151 | 2 | S70788 | curlin protein csg |
| 4 | 487.5 | 62.9 | 152 | 2 | D90806 | curlin major subun |
| 5 | 487.5 | 62.9 | 152 | 2 | H85665 | hypothetical prote |
| 6 | 107 | 13.8 | 2174 | 2 | E95985 | hypothetical Glyci |
| 7 | 97 | 12.5 | 1748 | 2 | S42136 | chjB protein - fet |
| 8 | 94.5 | 12.2 | 151 | 2 | JC6040 | fibrin protein ag |
| 9 | 94 | 12.1 | 645 | 2 | F70825 | probable PPE prote |
| 10 | 91.5 | 11.8 | 602 | 1 | PL0221 | leishmanolysin (EC |
| 11 | 91 | 11.7 | 573 | 2 | C86266 | Fif19.21 protein - |
| 12 | 91 | 11.7 | 615 | 2 | E70663 | probable PPE prote |
| 13 | 89.5 | 11.5 | 151 | 2 | S70787 | curlin nucleator p |
| 14 | 89.5 | 11.5 | 151 | 2 | C90806 | minor curlin subun |
| 15 | 89.5 | 11.5 | 151 | 2 | G85665 | curlin minor chain |
| 16 | 89.5 | 11.5 | 1322 | 2 | S7053 | ice nucleation pro |
| 17 | 89.5 | 11.5 | 1567 | 2 | S11672 | ice nucleation pro |
| 18 | 89 | 11.5 | 354 | 2 | E70663 | probable PPE prote |
| 19 | 88 | 11.4 | 407 | 2 | T21956 | hypothetical prote |
| 20 | 87.5 | 11.3 | 447 | 2 | G84687 | probable disease r |
| 21 | 87.5 | 11.3 | 1034 | 2 | JC2143 | ice nucleation act |
| 22 | 87.5 | 11.3 | 1258 | 2 | JQ0188 | ice nucleation pro |
| 23 | 87 | 11.2 | 590 | 2 | E70946 | probable PPE prote |
| 24 | 86.5 | 11.2 | 145 | 2 | AD3143 | conserved hypothet |
| 25 | 86.5 | 11.2 | 145 | 2 | H98144 | hypothetical prote |
| 26 | 86.5 | 11.2 | 151 | 2 | AH0635 | nucleation compone |
| 27 | 86 | 11.1 | 590 | 1 | A45621 | leishmanolysin (EC |
| 28 | 85.5 | 11.0 | 1053 | 2 | B70987 | probable PPE prote |
| 29 | 85.5 | 11.0 | 2204 | 2 | A70524 | probable PPE prote |

| | | | | | | |
|----|------|------|------|---|--------|--------------------|
| 30 | 85 | 11.0 | 1052 | 2 | AF2959 | conserved hypothet |
| 31 | 85 | 11.0 | 1341 | 2 | H98323 | hypothetical prote |
| 32 | 84.5 | 10.9 | 646 | 1 | S15915 | leishmanolysin (EC |
| 33 | 84.5 | 10.9 | 928 | 2 | C81265 | probable lipoprote |
| 34 | 84.5 | 10.9 | 978 | 2 | D81411 | probable lipoprote |
| 35 | 84 | 10.8 | 586 | 2 | T26667 | hypothetical prote |
| 36 | 84 | 10.8 | 599 | 2 | B42049 | leishmanolysin (EC |
| 37 | 84 | 10.8 | 599 | 2 | A44951 | leishmanolysin (EC |
| 38 | 84 | 10.8 | 1608 | 2 | A28182 | hemolysin A - Ser |
| 39 | 84 | 10.8 | 3300 | 2 | D70575 | probable PPE prote |
| 40 | 83.5 | 10.8 | 582 | 2 | F70675 | probable PPE prote |
| 41 | 83 | 10.7 | 277 | 2 | AB1390 | hypothetical cell |
| 42 | 83 | 10.7 | 343 | 2 | T05221 | hypothetical prote |
| 43 | 82.5 | 10.6 | 770 | 2 | T51024 | related to C2H2 z1 |
| 44 | 82.5 | 10.6 | 940 | 2 | D89723 | protein F39D8.1b (|
| 45 | 82.5 | 10.6 | 945 | 2 | T21998 | hypothetical prote |

ALIGNMENTS

RESULT 1

JC6039
fibrin protein agfa precursor - Salmonella enteritidis

C;Species: Salmonella enteritidis

C;Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999

C;Accession: JC6039; PC6015; A44898

R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A;Reference number: JC6039; MUID:96146512; PMID:8550497

A;Accession: JC6039

A;Molecule type: DNA

A;Residues: 1-151 <COL>

A;Cross-references: GB:U43280; NID:g1184712; PIDN:AA043599.1; PID:g1184714

A;Accession: PC6015

A;Molecule type: protein

A;Residues: 21-52 <CO2>

A;Experimental source: strain 27655-3b

A;Note: the authors translated the codon ACG for residue 44 as Ile

R;Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A;Title: Purification and characterization of thin, aggregative fimbriae from Salmonell

A;Reference number: A44898; MUID:91310586; PMID:1677357

A;Contents: 27655

A;Accession: A44898

A;Status: preliminary

A;Molecule type: protein

A;Residues: 21-33 <CO3>

A;Note: sequence extracted from NCBI backbone (NCBIP:45936)

C;Genetics:

A;Gene: agfa

C;Function:

A;Description: major component of thin aggregative fimbriae

A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C;Keywords: fimbria

F;1-20/Domain: signal sequence #status predicted <SIG>

F;21-151/Product: fibrin protein agfa #status experimental <VAT>

Query Match 88.1%; Score 683; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 1.4e-51;

Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGNNGSSGPDSTLSIYQVGSANAALAQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGNNGSSGPDSTLSIYQVGSANAALAQ 60

QY 61 SDARKSETTITSGYNGADYDQLVTRVVTHEMAFAFRNNATIDQWNAKNSDITVQYGG 120

Db 61 SDARKSETTITSGYNGADYDQLVTRVVTHEMAFAFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNALVNOTASDSSVMVRQVFGNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AI0635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; T. Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AI0635
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 88.1%; Score 683; DB 2; Length 151;
Best Local Similarity 90.1%; Pred. No. 1.4e-51;
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPSTLSIYYQGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPSTLSIYYQGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQYGG 120
DB 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S31202; S34560; S34559
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and CsgA
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70788
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1797265; PIDN:AACT4126.1; PID:g1787279;
A:Experimental source: strain K-12, substrain MGL655
R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6, 'V', 8-151 <OLS1>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42; 44-50 <OLS2>
R:Olsen, A.N.; Arngvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133, 'RQRSGMW' <OLS3>
A:Cross-references: EMBL:L04979; NID:G290424; PIDN:AAA23616.1; PID:g290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-151/Product: curlin #status experimental <MAT>

Query Match 65.7%; Score 509; DB 2; Length 151;
Best Local Similarity 68.2%; Pred. No. 1.1e-36;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPSTLSIYYQGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPSTLSIYYQGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQYGG 120
DB 61 TDARNSDLTITQHGNGGADYDQGGSDSSIDLTRQFGNSATLDQWNGNSNTVTKQFGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 GNGAAYDQTASNSVNVTVQFGNNATAHQY 151

RESULT 4

D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS1420

Query Match 62.9%; Score 487.5; DB 2; Length 152;
Best Local Similarity 66.4%; Pred. No. 7.6e-35;
Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPSTLSIYYQGSANAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPSTLSIYYQGSANAALAL 60
QY 60 QSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVQYGG 119


```

QY      11  AIIVSGSALGVVPOWCGGNNH--GGNSGSPDSTLSIYQVGSNAALALQSDARKSETT 69
DB      294  AVTPFGNGNVI-----GNGGNGFGGGT-----GNANIGLGNVGDG----- 331
QY      70  ITQSGYCGAGADYDQLVTRVVTHEMAFRNATIDQWAKNSDITVGYGCGNNAALVNQT 129
DB      332  --NVGFGSSGN-----TFQFGTG-----NNNIGILGTSNQIGFGGLN 369
QY      130  ASDSVWVRQVFGGNATAN 149
DB      370  SGGSN-----IGFGNSGTGN 384

RESULT 10
PL0221
Leishmanolysin (EC 3.4.24.36) precursor [validated] - Leishmania major
N;Alternate names: promastigote surface proteinase; surface endopeptidase glycoprotein 9
C;Species: Leishmania major
C;Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text_change 15-Sep-2000
C;Accession: PL0221; A27598; A60648
R;Button, L.L.; McMaster, W.R.
J. Exp. Med. 171, 589, 1990
A;Reference number: PL0221
A;Contents: erratum
A;Accession: PL0221
A:Molecule type: DNA
A;Residues: 1-602 <BUT>
A;Cross-references: GB|Y00647; NID:95554; PIDN:CAA68673.1; PID:99555
A;Note: this is a revision to the sequence from reference A27598
R;Button, L.L.; McMaster, W.R.
J. Exp. Med. 167, 724-729, 1988
A;Title: Molecular cloning of the major surface antigen of Leishmania.
A;Reference number: A27598; MUID:88154764; PMID:3346825
A;Accession: A27598
A;Status: significant sequence differences
A:Molecule type: DNA
R;Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.
Mol. Biochem. Parasitol. 37, 235-246, 1989
A;Title: Characterization of the promastigote surface protease of Leishmania as a member
A;Reference number: A60648; MUID:90114330; PMID:2608099
A;Accession: A60648
A:Molecule type: protein
A;Residues: 101,E,103-118,'SV',121-123 <BOU>
A;Experimental source: strain LEM513
R;Schlagenhauf, E.; Etges, R.; Metcalf, P.
submitted to the Brookhaven Protein Data Bank, March 1997
A;Reference number: A68135; PDB:1LML
A;Contents: annotation: X-ray crystallography, 1.86 angstroms, residues 100-407,412-498
A;Note: strain LRC-L119
C;Complex: homodimer
C;Function:
A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residues
A;Note: the activated form can activate the proenzyme form
C;Superfamily: leishmanolysin
C;Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; lip
F;1-39/Domain: signal sequence #status predicted <SIG>
F;40-100/Domain: activation peptide #status predicted <ATP>
F;101-577/Product: leishmanolysin #status experimental <MAR>
F;578-602/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;48,264,268,334/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
F;100-101/Cleavage site: Val-Val (autoolytic) #status experimental
F;125-142,191-230,314-386,393-455,406-425,415-489,466-510,515-565,535-558/Disulfide bond
F;264,268,334/Binding site: zinc, catalytic (His) (active) #status experimental
F;265/Active site: Glu #status predicted
F;300,407/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;577/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form
Query Match 11.8%; Score 91.5; DB 1; Length 602;
Best Local Similarity 41.7%; Pred. No. 3;
Matches 25; Conservative 2; Mismatches 12; Indels 21; Gaps 2;
QY      81  YDQLVTRVVTHEMAHAF-----RNNATIDQWAKNSDITV-----GOYG 119

```


Db 254 YDQLVTRVVTHEMAHALGFGSPFFEDARIIVANVNVNRKGFDFVPVINSSTAVAKARQYG 313

RESULT 11

C86266

C3P19_21 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001

C:Accession: C86266

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; anser, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C86266

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-573 <STO>

A:Cross-references: GB:AE005172; NID:G4850402; PIDN:AAD31072.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 11.7%; Score 91; DB 2; Length 573;

Best Local Similarity 26.9%; Pred. No. 3.2;

Matches 36; Conservative 19; Mismatches 49; Indels 30; Gaps 7;

QY 13 VVSGALAGVVPWGCGGNHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQ 72

Db 84 IVSGGTVEG---KYRNDGNG---ISGPDRTSDVYPQASFGAKGLNIDIOSNK--IAQ 135

QY 73 SGYGNAGDYDQLVTRVVTHEMAHAFRNA-TIDQWAKNS-----DITVQYQVGN 122

Db 136 QG-----STTWLNN--HGFGNANVPEMPVHNSYGAPPQGAQIPVQMSVNP 183

QY 123 AALVNOTASDSVM 136

Db 184 NVMNKSPSTQSFV 197

RESULT 12

E70663

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003

C:Accession: E70663

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitthead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70663

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-615 <COI>

A:Cross-references: GB:Z83860; GB:AL123456; NID:g3261681; PIDN:CAB06165.1; PID:e290763;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 11.7%; Score 91; DB 2; Length 615;

Best Local Similarity 22.5%; Pred. No. 3.4;

Matches 31; Conservative 14; Mismatches 47; Indels 46; Gaps 5;

QY 15 SGSALAGVVPWGCGGNHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQSG 74

Db 365 SGTGNIG---FGSGNNNIGFNSG-DGNIGFFNSGDGN-----TG 401

QY 75 YNGADYDQLVTRVVTHEMAHAFRNAATIDQWAKNSDITVQYQVGNNAALV---NOTAS 131

Db 402 FGNA-----GNINTGFVAGNLNTFGSAGNGVIFDGGNSGG 441

QY 132 DSSVMVRQVQFGNNATAN 149

Db 442 SFNVGFQNTGFGNSGAGN 459

RESULT 13

S70787

curlin nucleator protein csGB precursor - Escherichia coli (strain K-12)

N:Alternate names: csGB protein; curlin nucleation component; minor curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002

C:Accession: S70787; F64846

R:Hammar, M.; Arnyqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csG operons is required for production of fibronectin- and C

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70787

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C. A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4125.1; PID:g1787278

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csGB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tr

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 11.5%; Score 89.5; DB 2; Length 151;

Best Local Similarity 28.1%; Pred. No. 0.92;

Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSTLSIYQVGSANAALQSDARKSETTITQSGYGNAGDYDQLVTRVVTHEMAHAF 97

Db 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGAGTNNSQLRGGSKLLAVQAQGS 76

QY 98 RNNATIDQWAKNSDITVQYQVGNNAALVNOTASDSSVMVRQVQFGNNATANQY 151

Db 77 SNRAKIDQTDYNL-AYIDQAGSANDASISQAGVGNNTAMIIQKSGNGKANITQY 129

RESULT 14

C90806

minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subs

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: C90806

R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds

(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapcp 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|-------------|
| 1 | 683 | 88.1 | 151 | 1 | CSGA_SALTY |
| 2 | 509 | 65.7 | 151 | 1 | CSGA_ECOLI |
| 3 | 487.5 | 62.9 | 152 | 1 | CSGA_ECO57 |
| 4 | 94.5 | 12.2 | 151 | 1 | CSGB_SALTY |
| 5 | 91.5 | 11.8 | 602 | 1 | GP63_LEIMA |
| 6 | 89.5 | 11.5 | 151 | 1 | CSGB_ECOLI |
| 7 | 89.5 | 11.5 | 1322 | 1 | ICBA_PANAN |
| 8 | 89.5 | 11.5 | 1567 | 1 | ICEN_XANCT |
| 9 | 87.5 | 11.3 | 1034 | 1 | ICEN_PANAN |
| 10 | 87.5 | 11.3 | 1258 | 1 | ICEN_ERWHE |
| 11 | 86.5 | 11.2 | 151 | 1 | CSGB_SALTY |
| 12 | 86 | 11.1 | 590 | 1 | GP63_LEIDO |
| 13 | 84.5 | 10.9 | 646 | 1 | GP63_LEIME |
| 14 | 84 | 10.8 | 599 | 1 | GP63_LEICH |
| 15 | 84 | 10.8 | 1608 | 1 | HLVA_SERMA |
| 16 | 83.5 | 10.8 | 856 | 1 | AE7_AZOVI |
| 17 | 83 | 10.7 | 1656 | 1 | OMPB_RICVA |
| 18 | 81.5 | 10.5 | 392 | 1 | HMB1_HUMAN |
| 19 | 81.5 | 10.5 | 697 | 1 | SIM_DROME |
| 20 | 81.5 | 10.5 | 1028 | 1 | OVO_DROME |
| 21 | 80 | 10.3 | 331 | 1 | OMB2_NEIMB |
| 22 | 80 | 10.3 | 760 | 1 | YB1L_ECOLI |
| 23 | 79.5 | 10.3 | 475 | 1 | ABPB_STRCO |
| 24 | 79.5 | 10.3 | 548 | 1 | CEAK_ECOLI |
| 25 | 79.5 | 10.3 | 1210 | 1 | ICEN_FSEFL |
| 26 | 79 | 10.2 | 678 | 1 | YF48_MYCTU |
| 27 | 79 | 10.2 | 823 | 1 | NSP1_YEAST |
| 28 | 78.5 | 10.1 | 232 | 1 | DHN3_PEA |
| 29 | 78.5 | 10.1 | 641 | 1 | IND_APTGO |
| 30 | 77.5 | 10.0 | 1063 | 1 | SPF5_YEAST |
| 31 | 77.5 | 10.0 | 1148 | 1 | ICBK_FSESK |
| 32 | 77.5 | 10.0 | 1185 | 1 | MAPX_DROME |
| 33 | 77.5 | 10.0 | 1196 | 1 | ICVY_FSESK |

34 77 9.9 204 1 CORA_MEDSA
35 77 9.9 376 1 P33_MYCPE
36 77 9.9 439 1 ACMA_LACLA
37 76.5 9.9 487 1 Y442_MYCTU
38 76.5 9.9 881 1 YFCU_ECOLI
39 76 9.8 147 1 HFAA_CAUCR
40 76 9.8 362 1 P35_MYCPE
41 76 9.8 542 1 SCWE_YEAST
42 76 9.8 1571 1 C3G_DROME
43 75.5 9.7 163 1 HCY_NAIPH
44 75.5 9.7 394 1 OMSI_SALTY
45 75.5 9.7 424 1 COAA_BFPD

ALIGNMENTS

RESULT 1
CSGA_SALTY
ID CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RN Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RN Nature 413:848-852(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12844504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;

RT "Comparative Genomics of Salmonella enterica serovar Typhi strains Ty2
 RL and CT18.";
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 RN [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Searles G., Todd E.C.D.,
 Munto C.K., Kay C.M., Bansen P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for Salmonella species targeting agfA,
 the structural gene for thin, aggregative fimbriae.";
 RL J. Clin. Microbiol. 31:2263-2273(1993).
 RN [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AJ002301; CAA05317.1; -;
 DR EMBL; AF008749; AAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; A2016840; AAC069399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR FIC; JC6039; JC6039.
 DR StyGene; SGI0608; CSGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151
 FT CONFLICT 134 151
 FT REF. 6)
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 SQ
 Query Match 88.1%; Score 683; DB 1; Length 151;
 Best Local Similarity 90.1%; Pred. No. 1.3e-51;
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MKLLKVAAPAAIVSGSALAGVVPQGGGNGHNGSGSPDSTLSIIYQGSANAALQ 60
 DB 1 MKLLKVAAPAAIVSGSALAGVVPQGGGNGHNGSGSPDSTLSIIYQGSANAALQ 60
 QY 61 SPARKSETTITSGYXGNGADYDQLVTRVVVTHEMAHAFRNATIDOWNAKNSDITVQYGG 120
 DB 61 SPARKSETTITSGYXGNGADYDQLVTRVVVTHEMAHAFRNATIDOWNAKNSDITVQYGG 120
 QY 121 NNAALVNQTSDDSSVMVROVFGNNATANOY 151
 DB 121 NNAALVNQTSDDSSVMVROVFGNNATANOY 151

RESULT 2
 CSGA_ECOLI STANDARD; PRT; 151 AA.
 ID _CSGA_ECOLI AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=93211294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csgA, the subunit gene of fibronectin-binding curli in
 Escherichia coli.";
 RL Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two csg operons is required for production of
 fibronectin- and congo red-binding curli polymers in Escherichia coli
 K-12.";
 RL Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1232-1244(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / IMEL;
 RX MEDLINE=93033873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The Crl protein activates cryptic genes for curli formation and
 fibronectin binding in Escherichia coli HB101.";
 RL Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RX MEDLINE=91310586; PubMed=1677357; T.J., Kay W.W.;
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 Salmonella enteritidis.";
 RL J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 FIBRONECTIN.
 CC

```

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -.
DR EMBL; X90754; CAA62282.1; -.
DR EMBL; AE000205; AAC74126.1; -.
DR EMBL; D90741; BAA35832.1; -.
DR EMBL; D90742; BAA35840.1; -.
DR PIR; S70788; S70788.
DR EcoGene; EG11489; csGA.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.
FT CONFLICT 7 7 A -> E (IN REF. 1).
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;

Query Match 65.7%; Score 509; DB 1; Length 151;
Best Local Similarity 68.2%; Pred. No. 8.7e-37;
Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVGOYGG 120
DB 61 TDARNSDLTITQHGNGGADYVQGSDDSSIDLTRQFGNSATLDQWNGKSEMTVKQFGG 120

QY 121 NNAALVNQTSADSSVMVQVFGNNATANY 151
DB 121 GNGAAVDOTASNSVNVTVQVFGNNATAHQY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21074935; PubMed=11206551;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).

```

```

[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaishi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuda E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE275733; AAK53212.1; -.
DR EMBL; AE005315; AAG55788.1; -.
DR EMBL; AP002554; BAB34843.1; -.
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 62.9%; Score 487.5; DB 1; Length 152;
Best Local Similarity 66.4%; Pred. No. 6e-35;
Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALQ 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGDPSTLSIYQGSANAALQ 60
QY 60 QSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVGOYGG 119
DB 61 QADARNSDLTITQHGNGGADYVQGSDDSSIDLTRQFGNSATLDQWNGKSHMTVKQFG 120
QY 120 GNGAAVDOTASNSVNVTVQVFGNNATANY 151
DB 121 GNGAAVDOTASNSVNVTVQVFGNNATAHQY 152

RESULT 4
CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";

```



```
FT METAL 264 264 ZINC (CATALYTIC).
FT ACT_SITE 265 265
FT METAL 268 268 ZINC (CATALYTIC).
FT METAL 334 334 ZINC (CATALYTIC).
FT METAL 334 334
FT DISULFID 125 142
FT DISULFID 191 230
FT DISULFID 314 386
FT DISULFID 333 455
FT DISULFID 406 425
FT DISULFID 415 489
FT DISULFID 466 510
FT DISULFID 515 565
FT DISULFID 535 558
FT CARBOHYD 300 300
FT CARBOHYD 407 407
FT LIPID 577 577
FT STRAND 101 102
FT STRAND 107 108
FT STRAND 111 114
FT HELIX 116 119
FT TURN 121 122
FT TURN 128 129
FT STRAND 131 133
FT STRAND 139 141
FT HELIX 144 146
FT HELIX 150 158
FT TURN 159 159
FT HELIX 160 169
FT TURN 170 171
FT STRAND 172 174
FT STRAND 177 178
FT STRAND 180 181
FT TURN 189 190
FT HELIX 191 193
FT HELIX 198 202
FT TURN 203 203
FT STRAND 205 206
FT STRAND 210 215
FT TURN 221 222
FT STRAND 226 232
FT TURN 234 235
FT STRAND 238 244
FT HELIX 247 249
FT HELIX 256 269
FT TURN 270 271
FT HELIX 274 279
FT TURN 280 281
FT STRAND 283 286
FT HELIX 289 291
FT STRAND 296 299
FT HELIX 302 312
FT TURN 313 313
FT TURN 315 316
FT STRAND 320 322
FT TURN 328 332
FT STRAND 335 335
FT TURN 337 339
FT TURN 341 342
FT STRAND 343 343
FT TURN 344 345
FT STRAND 353 353
FT HELIX 356 364
FT TURN 365 366
FT STRAND 369 370
FT HELIX 372 374
FT TURN 380 383
FT HELIX 386 390
FT STRAND 394 395
FT TURN 396 397
FT STRAND 398 399
FT TURN 402 404
FT STRAND 413 414
FT TURN 417 418

ZINC (CATALYTIC).
ZINC (CATALYTIC).
ZINC (CATALYTIC).

N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
GPI-anchor amidated asparagine.

FT STRAND 421 425
FT STRAND 428 429
FT HELIX 435 437
FT TURN 443 444
FT STRAND 445 446
FT TURN 450 454
FT STRAND 458 465
FT TURN 466 467
FT HELIX 470 472
FT TURN 475 477
FT HELIX 478 480
FT TURN 485 486
FT STRAND 487 494
FT STRAND 496 496
FT STRAND 506 516
FT TURN 517 520
FT STRAND 521 525
FT TURN 527 528
FT STRAND 533 534
FT TURN 537 538
FT STRAND 540 542
FT HELIX 543 545
FT TURN 546 546
FT STRAND 550 550
FT TURN 552 553
FT STRAND 555 557
FT HELIX 561 565
FT TURN 566 567
FT HELIX 569 572
FT TURN 573 573
SQ SEQUENCE 602 AA; 63953 MW; 982BF3245D87C43E CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 602;
Best Local Similarity 41.7%; Pred. NO. 1.6;
Matches 25; Conservative 2; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDQLVTRVVTHEMAHAF-----RNNATIDOWNAKNSDITV-----GOYG 119
      |||||
Db 254 YDQLVTRVVTHEMAHALGFGPGFFEDARIVANPVNVRGNFDVPVINSSTAVAKAEQYG 313

RESULT 6
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
RT K-12.";
RL Mol. Microbiol. 18:661-670(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
```

```

Science 277:1453-1474 (1997).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:1137-1155 (1996).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Foster G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimallanta E.T., Pocomousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
[5]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
[6]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngqvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the esgA promoter in
Escherichia coli can be achieved in vivo by sigma 70 in the absence
of the nucleoid-associated protein H-NS.";
RL Mol. Microbiol. 13:1021-1032 (1994).
-1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI ARE
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
FIBRINECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CURLIN MONOMERS.
-1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X30754; CAA62281.1; -
EMBL; A800205; AAC74125.1; -
EMBL; D90741; BAA35831.1; -
EMBL; A8005315; AAG55787.1; -
EMBL; A8002554; BAB34842.1; -
PIR; C90806; C90806.
PIR; G85665; G85665.
PIR; S70787; S70787.
EcGene; EGI2621; cs9B.
FimBria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.

```

```

SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 151;
Best Local Similarity 28.1%; Pred. No. 0.5; Mismatches 62; Indels 5; Gaps 2;
Matches 32; Conservative 15;
QY 38 SSGPSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAF 97
Db 21 AGYDLANSEYNF----AVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQSGS 76
QY 98 RNNATIDQWNAKNSDITVGCYQYGNNAALVNQTASDSSVMVROVGFNNATANQY 151
Db 77 SNRAKIDQTDYNLAYIDQAGSANDASISQAYGNTAMITQKSGNKANITQY 129
RESULT 7
ICEA_PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9002494; PubMed=25990935;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
to those of Pseudomonas species and regions required for ice
nucleation activity.";
RL FEBS Lett. 258:297-300 (1989).
CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
crystallization in supercooled water.
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A
REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein
family.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X17316; CAA35194.1; -
PIR; S07053; S07053.
HSP; P06620; IINA.
InterPro; IPR000258; Ice_nucleatn.
Pfam; PF00818; Ice_nucleation; 69.
PRINTS; PR00327; ICENUCLEATN.
PROSITE; PS00314; ICE_NUCLEATION; 49.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1322 AA; 131094 MW; 89B0BE24AA837039 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 1322;
Best Local Similarity 26.8%; Pred. No. 5.9;
Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;
QY 34 GGNSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEM 93
Db 933 GSTSTAGPSSSL-IAGYGSTQTA-----GYNISLTAGYS----- 966
QY 94 AFAFRNNATIDQWNAKNSDITVGCYG-----GNNAALV-----NQTASDSSVMVROVGF 143

```



```

Db 967 -----TOTGENSEDLTTG-YGSTSTAGYESSLIAGYGSTQTASFXTLM--AGYG 1013
QY 144 NNATANQ 150
   : : : : :
Db 1014 SSQTARE 1020

RESULT 8
ICEN_XANCT
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X56S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RL Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC -!- crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52970; CAA37140.1; -.
CC HSSP; P06620; LINA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 81.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 57.
KW Ice nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 1567;
Best Local Similarity 26.4%; Pred. No. 7.2;
Matches 47; Conservative 23; Mismatches 59; Indels 49; Gaps 12;

QY 14 VSGSALAG-----VFPQWG--GGNHN-----GGNNSGPDSTLSIYQYGSNAALAL 59
Db 205 VYGSTLTGADOSRLVAGYGSTTETAGDHDLIAGYGSTGTAGSDSSI-LAGYGSTQTAGR 263
QY 60 Q-----SDARKSETTTTQSGYG-----NGAD-----YDQLVTRVVTHEMAHFRNAT 102
Db 264 STLTAGYGSTQTAGSGRLTSGYSTATSGSDSAVIGYSTGTAGSESSLTAGYGSTQT 323
QY 103 IDQWNAKNSDITVGYG-----GNNAALV-----NOTASDSSVMVROVGFNNATANQ 150
Db 324 A----RKGSIDITAG-YGSTGTAGSDSALIAGYGSTQTAGSESSLT--AGYGSTQTARK 374

```

```

RESULT 9
ICEN_PANAN
ID ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora.";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC -!- crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14992; BAA03636.1; -.
CC PIR; JC2143; JC2143.
CC HSSP; P06620; LINA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 51.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match 11.3%; Score 87.5; DB 1; Length 1034;
Best Local Similarity 25.2%; Pred. No. 6.6;
Matches 41; Conservative 18; Mismatches 41; Indels 63; Gaps 9;

QY 12 IVVSGSALAGVVPQW--GGGNHNGGNS-----GPDSTLSIYQYGSNAAL 57
Db 161 IATYGSTLTGTHQSOLIAGYGSTETAGDSSTLIAGYGSTGTAGSDSTL-VAGYGSTQTA- 218
QY 58 ALQSDARKSETTITQSGYCGNGADYDQLVTRVVTHEMAHFRNATIDQWNAKNSDITVQ 117
Db 219 -----GEESQWAGYS-----TQGMKGSDLTAG- 243
QY 118 YG-----GNNAALV-----NOTASDSSVMVROVGFNNATANQ 150
Db 244 YGSTGTAGDSSLIAGYGSTQTAGDSSLT--AGYGSTQTACK 284

RESULT 10
ICEN_ERWHE
ID ICEN_ERWHE STANDARD; PRT; 1258 AA.
AC P16239;

```

```

DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ICE nucleation protein.
GN ICEE.

OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; W26382; AAA24823.1; -.
CC PIR; JQ0188; JQ0188.
CC DR HSP; P06620; IINA.
CC DR InterPro; IPR000258; Ice nucleatn.
CC DR Pfam; PF00818; Ice nucleation; 65.
CC DR PRINTS; PR00327; ICNUCLEATN.
CC DR PROSITE; PS00314; ICE_NUCLEATION; 45.
CC Ice nucleation; Repeat; Outer membrane.
CC FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1258 AA; 125084 MW; 59088A130077FBD4 CRC64;

Query Match 11.3%; Score 87.5; DB 1; Length 1258;
Best Local Similarity 25.2%; Pred. No. 8.2; Indels 53; Gaps 9;
Matches 41; Conservative 18; Mismatches 41;

QY 12 IVVSGSALAGVVPQW--GGGNHNGGNS-----GPDSTLSIYQVGSANAAL 57
Db 161 IATYGSTLSGTHQSOLIAGYGSTAGDSSTLIAGYGSTGTAGADSTL-VAGYGSTQTA- 218
QY 58 ALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITV 117
Db 219 -----GEESQVAGYGS-----TGTGKSGDLTAG- 243
QY 118 YG-----GNNAALV-----NOTASDSSVMVRQVGFNNATANQ 150
Db 244 YGSTGTAGDSSLIAGYGSTQTAGEDSSLT--AGYGSTQTAQK 284

RESULT 11
CSGB SALT1
ID CSGB SALT1 STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR TI177.

OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahinia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanani V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627269; CAD08267.1; -.
CC DR EMBL; AB016840; AAC69400.1; -.
CC KW Fimbrin; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 11.2%; Score 86.5; DB 1; Length 151;
Best Local Similarity 26.2%; Pred. No. 0.89; Indels 37; Gaps 4;
Matches 34; Conservative 17; Mismatches 42;

QY 10 AAIWVGSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQVGSANAALQSDARKSETT 69
Db 58 ARVQSGSKLLSVISQ--EGENNRKAVQDQAGNYFAYIEQTGNAN-----DAS 103
QY 70 ITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVQYGGNNAALVQNT 129
Db 104 ISQAYNGSA-----AIIQKSGNKANIT--QYGTOKTAVVQVK 140
QY 130 ASDSSVMVRQ 139
Db 141 QSHMAIRVTQ 150

RESULT 12
GP63 LEIDO
ID GP63 LEIDO STANDARD; PRT; 590 AA.
AC P232Z3;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)

```



```

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 646 AA; 69054 MW; FE448DDC78C10B0A CRC64;

Query Match 10.9%; Score 84.5; DB 1; Length 646;
Best Local Similarity 36.5%; Pred. No. 6.9;
Matches 23; Conservative 7; Mismatches 12; Indels 21; Gaps 2;

QY 81 YDQLVTRVTVTHMAHAFRNAT-----IDQNAKNKSDITV-----GOYG 119
256 YDQLVTRVTVTHMAHAFRNATGFFGAGVGVQVPHLRKDFNVSVITSTVVAAREQYG 315
DB 120 GNN 122
316 CNS 318

RESULT 14
GP63_LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
CX NCBI_TaxID=44271;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=92112918; PubMed=1370484;
RX Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,
RA Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially
RT expressed during development of Leishmania donovani chagasi
RT Promastigotes to an infectious form.";
RL J. Biol. Chem. 267:1888-1895(1992).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M30672; AAA29238.1; -
CC EMBL; M28527; AAA29235.1; -
CC FIR; A44951; A44951.
CC HSSP; P08148; 1LMML.
CC MEROPS; M08.001; -
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001577; Peptidase_M8.

```

```

DR PFam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN NATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 BY SIMILARITY.
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 585 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By
FT similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;

Query Match 10.8%; Score 84; DB 1; Length 599;
Best Local Similarity 37.5%; Pred. No. 7;
Matches 24; Conservative 5; Mismatches 29; Indels 6; Gaps 1;

QY 39 SGPDSTLSIVQYGSANAALA-----LQSDARKSETTITQSGYNGGADYDQLVTRVTHE 92
203 SNTDFVMYVASVPSEGVLAWTTCQVFSGDGHPAVGVINIPRAANTASRYDQLVTRVTHE 262
DB 93 MAHA 96
263 MAHA 266

RESULT 15
HLXA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN SHLA.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Serratia.
CX NCBI_TaxID=615;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RX STRAIN=SN8;
RX MEDLINE=88257037; PubMed=3290200;
RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia
RT marcescens.";
RL J. Bacteriol. 170:3177-3188(1988).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
CC REQUIRES SHLB FUNCTION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPWA).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-28
Perfect score: 775
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....DSSVMVQVGFGNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mic:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp Vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 674 | 87.0 | 152 | 2 | O33802 |
| 2 | 566.5 | 73.1 | 150 | 2 | Q7X243 |
| 3 | 530 | 68.4 | 149 | 2 | Q7X240 |
| 4 | 484.5 | 62.5 | 152 | 16 | Q8CW63 |
| 5 | 411.5 | 53.1 | 150 | 2 | Q7X237 |
| 6 | 295 | 38.1 | 76 | 2 | O54069 |
| 7 | 122 | 15.7 | 29 | 2 | Q8S335 |
| 8 | 112 | 14.5 | 130 | 16 | Q89J14 |
| 9 | 111.5 | 14.4 | 3501 | 16 | Q8Y106 |
| 10 | 111.5 | 14.4 | 3552 | 16 | Q8XSD6 |
| 11 | 110 | 14.2 | 139 | 16 | Q8EIH3 |
| 12 | 108 | 13.9 | 1422 | 16 | Q8EFU3 |
| 13 | 107 | 13.8 | 2174 | 16 | Q92JU8 |
| 14 | 106.5 | 13.7 | 502 | 16 | Q8EIH4 |
| 15 | 99.5 | 12.8 | 171 | 16 | Q89J13 |
| 16 | 99 | 12.8 | 714 | 16 | Q7U5X6 |

| | | | | | |
|----|------|------|------|----|--------|
| 17 | 98 | 12.6 | 179 | 2 | O33801 |
| 18 | 97.5 | 12.6 | 154 | 16 | Q89J15 |
| 19 | 97 | 12.5 | 1748 | 5 | Q84821 |
| 20 | 95 | 12.3 | 368 | 16 | Q8EW56 |
| 21 | 95 | 12.3 | 400 | 10 | Q8L4R8 |
| 22 | 95 | 12.3 | 1410 | 16 | Q8CWJ0 |
| 23 | 94.5 | 12.2 | 160 | 16 | Q8CW64 |
| 24 | 94 | 12.1 | 362 | 16 | Q8EV84 |
| 25 | 94 | 12.1 | 645 | 16 | Q7U1C5 |
| 26 | 94 | 12.1 | 646 | 16 | O53818 |
| 27 | 93.5 | 12.1 | 644 | 5 | Q8MY19 |
| 28 | 93 | 12.0 | 157 | 16 | Q88HG0 |
| 29 | 92.5 | 11.9 | 348 | 13 | Q93397 |
| 30 | 92.5 | 11.9 | 644 | 5 | O43994 |
| 31 | 92.5 | 11.9 | 1765 | 16 | Q7V885 |
| 32 | 92 | 11.9 | 598 | 5 | Q25275 |
| 33 | 91 | 11.7 | 152 | 2 | Q7X241 |
| 34 | 91 | 11.7 | 573 | 10 | Q3SAF2 |
| 35 | 91 | 11.7 | 614 | 16 | Q7TYR8 |
| 36 | 91 | 11.7 | 615 | 16 | P95249 |
| 37 | 91 | 11.7 | 1408 | 16 | Q8E833 |
| 38 | 90.5 | 11.7 | 1286 | 2 | Q841Y5 |
| 39 | 89.5 | 11.5 | 151 | 2 | Q7X238 |
| 40 | 89.5 | 11.5 | 151 | 16 | Q7UC21 |
| 41 | 89.5 | 11.5 | 153 | 16 | Q89J16 |
| 42 | 89.5 | 11.5 | 160 | 16 | Q83RU7 |
| 43 | 89.5 | 11.5 | 1333 | 16 | Q8PD38 |
| 44 | 89.5 | 11.5 | 2411 | 16 | Q8PF72 |
| 45 | 89 | 11.5 | 354 | 16 | P95248 |

ALIGNMENTS

RESULT 1

O33802
ID O33802 PRELIMINARY; PRT; 152 AA.
AC O33802; (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JAN-1998 (TREMBLrel. 19, Last annotation update)
DE Agfa protein (Fragment).
GN AGFA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells."
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CAA04151.1; -
FT NON_TER 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.0%; Score 674; DB 2; Length 152;
Best Local Similarity 88.7%; Pred. No. 1.1e-47;
Matches 134; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

| | | |
|----|-----|---|
| QY | 1 | MKLLKVAAPAAIVVSGSALAGVVVPGGGGNGGNSGGPDSTLSIYGYGANAALAIQ 60 |
| Db | 1 | MKLLKVAAPAAIVVSGSALAGVVVPGGGGNGGNSGGPDSTLSIYGYGANAALAIQ 60 |
| QY | 61 | SDARKSETTITGSGYNGGADYDQLVTRVVVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120 |
| Db | 61 | SDARKSETTITGSGYNGGADYDQLVTRVVVTHEMAHAFRNATIDQWNAKNSDITVQYGG 120 |
| QY | 121 | NNAALVNTASDSSVNVVQVGFGNATANQY 151 |

```

Db 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
|||||
RESULT 2
Q7X243 ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 73.1%; Score 566.5; DB 2; Length 150;
Best Local Similarity 76.8%; Pred. No. 6.5e-39;
Matches 116; Conservative 10; Mismatches 24; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPGDSTLSIYQYGSANAALQ 60
|||||
Db 1 MKLLQVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPGDSTLSIYQYGSANAALQ 59
|||||
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVGOYGG 120
|||||
Db 60 SDARKSDTTIHQNGFGNGADVQGGSDNSTDLTQNGFGNNATIDQWNGKNSDITVSQYGG 119
|||||
QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVWVRQVGFNNATANQY 150
|||||

RESULT 3
Q7X240 ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 68.4%; Score 530; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 6.3e-36;
Matches 108; Conservative 16; Mismatches 25; Indels 2; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPGDSTLSIYQYGSANAALQ 60
|||||

```

```

Db 1 MKLLKVAFAAIVVSGSALAGVVPQW--GGNHGGSGSYGPDSSLSIYQYGSNNSANAQ 58
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVGOYGG 120
|||||
Db 59 SDARKSDVTITQHGNGAVVGQGADDSITSLKQTGFQNSATIDQWNAKNAIDSVTQFGG 118
|||||
QY 121 NNAALVNQTASDSSVWVRQVGFNNATANQY 151
|||||
Db 119 RINGALVNQTASDSSVWVRQVGFNNATANQY 149
|||||

RESULT 4
Q8CW63 ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016759; AAN79779.1; -
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 62.5%; Score 484.5; DB 16; Length 152;
Best Local Similarity 66.4%; Pred. No. 3.4e-32;
Matches 101; Conservative 15; Mismatches 35; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW--GGGNGHGGNSGPGDSTLSIYQYGSANAALQ 59
|||||
Db 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGGNGHGGNSGPNSELNIYQYGGNSALAQ 60
|||||
QY 60 QSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVGOYGG 119
|||||
Db 61 QADARNSLTITQHGNGADVQGGSDNSTDLTQNGFGNSATLDQWNGKSDTMTVQFG 120
|||||
QY 120 GNNALVNQTASDSSVWVRQVGFNNATANQY 151
|||||
Db 121 GNGHADVQTASDSSVWVRQVGFNNATANQY 152
|||||

RESULT 5
Q7X237 ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;

```



```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5DBB2D872DF15F3 CRC64;

Query Match 53.1%; Score 411.5; DB 2; Length 150;
Best Local Similarity 57.6%; Pred. No. 3.1e-26;
Matches 87; Conservative 25; Mismatches 38; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKFIKVAALAAIVVSGSAGMAGMIQ-GWGCHGHGGYGGPNSLTNIYQNGGNSALALQ 59
QY 61 SDARKSETTTTQSYGNGADYDQLVTRVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 TDAENSVLINSQTGNGADYGGQSDSSINLTQNGFGNSATLDQWNSKDSVMNVSYQGG 119
QY 121 NNAALVNQTSDDSSVMVROVGFGNATANOY 151
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 LNALVDQTSNSTVNVNTQIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON TER 1
FT NON TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.1%; Score 295; DB 2; Length 76;
Best Local Similarity 78.9%; Pred. No. 4.7e-17;
Matches 60; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 30 GNHGGNGSGSPDSTLSIYQYGSANAALALQSDARKSETTTTQSYGNGADYDQLVTRV 89
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GNHGGNGSGSPDSTLSIYQYGSANAALALQSDARKSETTTTQSYGNGADYDQGDNST 60
QY 90 THEMAHAFRNATIDQ 105
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IELTQNGFRNATIDQ 76

RESULT 7
Q54069 PRELIMINARY; PRT; 29 AA.
AC Q54069
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curlation of Escherichia coli O78:K80 isolates associated with
RT IS1 insertion in csgB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON TER 29
FT NON TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0021;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAALAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q89JI4 PRELIMINARY; PRT; 130 AA.
AC Q89JI4
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BLI5299 protein.
GN BLI5299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasanoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP05954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.5%; Score 112; DB 16; Length 130;
Best Local Similarity 26.5%; Pred. No. 0.085;
Matches 40; Conservative 28; Mismatches 55; Indels 28; Gaps 6;

QY 4 LKVAAPAAIVVSGSALAGVVPQWGGNGHNGSGSPDSTLSIYQYGSANAALALQSDA 63
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MRITYLVATAIALSALTIVDAQ-----AGNSA-----SVLQFGTNSFSQTGS 45
QY 64 RKSETTTTQSYGNGADYDQLVTRVTHEMAHAFRNATIDQW----NAKNSDITVGOY 119
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 46 TSNNATTLQFATNTA-----TLTQGSLLTV--NTAVTGGGTATATNTALT-GQVG 96

Query Match 14.5%; Score 112; DB 16; Length 130;
Best Local Similarity 26.5%; Pred. No. 0.085;
Matches 40; Conservative 28; Mismatches 55; Indels 28; Gaps 6;

QY 120 GNNALVNQTSDDSSVMVROVGFGNATANOY 150
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 97 GNSSLIGIGANNTAGVGLGILNGSTILQ 127

RESULT 9
Q8Y106 PRELIMINARY; PRT; 3501 AA.
ID Q8Y106
AC Q8Y106
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

```

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DR Probable hemagglutinin-related protein.
 GN RSC0887 OR RS061116.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RL Nature 415:497-502(2002).
 DR EMBL: AL646061; CAD14589.1; -.
 DR GO: GO:0004519; F:endonuclease activity; IEA.
 DR GO: GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro: IPR001604; Endonuclease.
 DR InterPro: IPR008638; Haemagg_act.
 DR Pfam: PF05594; Fil_haemagg; 20.
 DR Pfam: PF05960; Haemagg_act; 1.
 DR PROSITE: PS01070; NUCLEASE_NON_SPEC; 1.
 KW Complete proteome.
 SQ SEQUENCE 3501 AA; 348421 MW; 290841C99018A107 CRC64;

 Query Match 14.4%; Score 111.5; DB 16; Length 3501;
 Best Local Similarity 29.3%; Pred. No. 5.2;
 Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

 QY 15 SCSALAGVVPQWGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
 DB 2431 SCSHFSTAGPWSGDLGRNVGGGPNSSG---VGLAPYGSASHSADNAAGNSRQNASVVG 2486
 QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVG 116
 DB 2487 KSVQVQARTGDTIVSGSGISALSDVLLAKQKQVDIVA---GNDTSSR-HEHSDRTIG 2541
 QY 117 QYGGNNAALVNQTASDSSVM 136
 DB 2542 DLGGNGYSGTVGVSASSTL 2561

 RESULT 10
 QXASD6 PRELIMINARY; PRT; 3552 AA.
 ID Q8XSD6
 AC Q8XSD6;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE Probable hemagglutinin-related protein.
 GN RSP0540 OR RS06117.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RL Nature 415:497-502(2002).
 DR EMBL: AL646061; CAD14589.1; -.
 DR GO: GO:0004519; F:endonuclease activity; IEA.
 DR GO: GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro: IPR001604; Endonuclease.
 DR InterPro: IPR008638; Haemagg_act.
 DR Pfam: PF05594; Fil_haemagg; 20.
 DR Pfam: PF05960; Haemagg_act; 1.
 DR PROSITE: PS01070; NUCLEASE_NON_SPEC; 1.
 KW Complete proteome.
 SQ SEQUENCE 3501 AA; 348421 MW; 290841C99018A107 CRC64;

 Query Match 14.4%; Score 111.5; DB 16; Length 3501;
 Best Local Similarity 29.3%; Pred. No. 5.2;
 Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

 QY 15 SCSALAGVVPQWGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
 DB 2431 SCSHFSTAGPWSGDLGRNVGGGPNSSG---VGLAPYGSASHSADNAAGNSRQNASVVG 2486
 QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVG 116
 DB 2487 KSVQVQARTGDTIVSGSGISALSDVLLAKQKQVDIVA---GNDTSSR-HEHSDRTIG 2541
 QY 117 QYGGNNAALVNQTASDSSVM 136
 DB 2542 DLGGNGYSGTVGVSASSTL 2561

 RESULT 10
 QXSIH3 PRELIMINARY; PRT; 139 AA.
 ID Q8EIH3
 AC Q8EIH3;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DE Minor curlin subunit CsgB, putative.
 GN S00866.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer R.T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL: AE015532; AAN53942.1; -.
 DR TIGR: S00866;
 KW Complete proteome.
 SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

 Query Match 14.2%; Score 110; DB 16; Length 139;
 Best Local Similarity 28.3%; Pred. No. 0.13;
 Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 3;

 QY 39 SGPDSSTLSIYQGSANALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFN 98
 DB 41 SGRDNLIDLVOQGTANQGVFQSGSDNS-AVYQAGNDN---ISLVTQIGT----- 87
 QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVQVGFN 144

RL Nature 415:497-502(2002).
 DR EMBL: AL646079; CAD17691.1; -.
 DR GO: GO:0046821; C:extrachromosomal DNA; IEA.
 DR GO: GO:0004519; F:endonuclease activity; IEA.
 DR GO: GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro: IPR001604; Endonuclease.
 DR InterPro: IPR008638; Haemagg_act.
 DR Pfam: PF05594; Fil_haemagg; 20.
 DR Pfam: PF05960; Haemagg_act; 1.
 DR PROSITE: PS01070; NUCLEASE_NON_SPEC; 1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 3552 AA; 352934 MW; C5432AABE2CCF59C CRC64;

 Query Match 14.4%; Score 111.5; DB 16; Length 3552;
 Best Local Similarity 29.3%; Pred. No. 5.3;
 Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;

 QY 15 SCSALAGVVPQWGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
 DB 2430 SCSHFSTAGPWSGDLGRNVGGGPNSSG---VGLAPYGSASHSADNAAGNSRQNASVVG 2485
 QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFNNATIDQWNAKNSDITVG 116
 DB 2486 KSVQVQARTGDTIVSGSGISALSDVLLAKQKQVDIVA---GNDTSSR-HEHSDRTIG 2540
 QY 117 QYGGNNAALVNQTASDSSVM 136
 DB 2541 DLGGNGYSGTVGVSASSTL 2560

 RESULT 11
 Q8EIH3 PRELIMINARY; PRT; 139 AA.
 ID Q8EIH3
 AC Q8EIH3;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DE Minor curlin subunit CsgB, putative.
 GN S00866.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.B., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer R.T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser C.M.,
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL: AE015532; AAN53942.1; -.
 DR TIGR: S00866;
 KW Complete proteome.
 SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

 Query Match 14.2%; Score 110; DB 16; Length 139;
 Best Local Similarity 28.3%; Pred. No. 0.13;
 Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 3;

 QY 39 SGPDSSTLSIYQGSANALQSDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFN 98
 DB 41 SGRDNLIDLVOQGTANQGVFQSGSDNS-AVYQAGNDN---ISLVTQIGT----- 87
 QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVQVGFN 144

Query Match 13.7%; Score 106.5; DB 16; Length 502;
Best Local Similarity 24.2%; Pred. No. 1.2;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds
(without alignments)
950.215 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAAPRAIVVSSALA.....DSSVMYRVQVFGNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 768 | 100.0 | 151 | 3 | AAB36355 | Aab36355 Agfa::PT3 |
| 2 | 700 | 91.1 | 151 | 3 | AAB36352 | Aab36352 Agfa::PT3 |
| 3 | 681 | 88.7 | 151 | 2 | AAR74625 | Aar74625 Agfa sequ |
| 4 | 681 | 88.7 | 151 | 3 | AAB36341 | Aab36341 Salmonell |
| 5 | 676 | 88.0 | 151 | 2 | AAW23570 | Aaw23570 Salmonell |
| 6 | 659 | 85.8 | 151 | 3 | AAB36347 | Aab36347 Agfa::PT3 |
| 7 | 657 | 85.5 | 151 | 3 | AAB36351 | Aab36351 Agfa::PT3 |
| 8 | 613 | 79.8 | 151 | 3 | AAB36354 | Aab36354 Agfa::PT3 |
| 9 | 603 | 78.5 | 151 | 3 | AAB36346 | Aab36346 Agfa::PT3 |
| 10 | 600 | 78.1 | 151 | 3 | AAB36353 | Aab36353 Agfa::PT3 |
| 11 | 599 | 78.0 | 151 | 3 | AAB36349 | Aab36349 Agfa::PT3 |
| 12 | 597 | 77.7 | 151 | 3 | AAB36350 | Aab36350 Agfa::PT3 |
| 13 | 566 | 73.7 | 151 | 3 | AAB36348 | Aab36348 Agfa::PT3 |
| 14 | 517 | 67.3 | 151 | 3 | AAB36343 | Aab36343 Escherich |
| 15 | 512 | 66.7 | 151 | 7 | ABR82651 | AbR82651 E. coli C |
| 16 | 496 | 64.6 | 120 | 2 | AAR62761 | Aar62761 Agfa sequ |
| 17 | 496 | 64.6 | 120 | 2 | AAW23569 | Aaw23569 Salmonell |
| 18 | 439 | 57.2 | 142 | 2 | AAR52664 | Aar52664 Fibronect |
| 19 | 367 | 47.8 | 122 | 2 | AAR52663 | Aar52663 FNB curli |
| 20 | 215 | 28.0 | 45 | 3 | AAB36316 | Aab36316 Salmonell |
| 21 | 132 | 17.0 | 22 | 3 | AAB36318 | Aab36318 Salmonell |
| 22 | 115 | 15.0 | 22 | 3 | AAB36325 | Aab36325 Salmonell |
| 23 | 115 | 15.0 | 22 | 3 | AAB36339 | Aab36339 Salmonell |
| 24 | 115 | 15.0 | 22 | 3 | AAB36320 | Aab36320 Salmonell |
| 25 | 113 | 14.7 | 24 | 7 | ABR82644 | AbR82644 E. coli c |

| | | | | | | |
|----|-------|------|------|---|----------|--------------------|
| 26 | 109 | 14.2 | 23 | 3 | AAB36340 | Aab36340 Salmonell |
| 27 | 109 | 14.2 | 23 | 3 | AAB36324 | Aab36324 Salmonell |
| 28 | 109 | 14.2 | 23 | 3 | AAB36319 | Aab36319 Salmonell |
| 29 | 104.5 | 13.6 | 151 | 3 | AAB36344 | Aab36344 Escherich |
| 30 | 102 | 13.3 | 26 | 7 | ABR82649 | AbR82649 E. coli V |
| 31 | 101.5 | 13.2 | 520 | 6 | AAO16497 | Aao16497 Argiope t |
| 32 | 99.5 | 13.0 | 151 | 3 | AAB36342 | Aab36342 Salmonell |
| 33 | 98 | 12.8 | 26 | 7 | ABR82645 | AbR82645 E. coli c |
| 34 | 96 | 12.5 | 19 | 3 | AAB36323 | Aab36323 Salmonell |
| 35 | 96 | 12.5 | 19 | 3 | AAB36336 | Aab36336 Salmonell |
| 36 | 96 | 12.5 | 19 | 3 | AAB36328 | Aab36328 Salmonell |
| 37 | 95.5 | 12.4 | 908 | 4 | ABR65007 | AbR65007 Drosophil |
| 38 | 95 | 12.4 | 718 | 3 | AAY78364 | Aay78364 H. pylori |
| 39 | 94.5 | 12.3 | 908 | 4 | ABR62355 | AbR62355 Drosophil |
| 40 | 94 | 12.2 | 354 | 7 | ABO23520 | AbO23520 Mycobacte |
| 41 | 93.5 | 12.2 | 2204 | 6 | ABU36440 | Abu36440 Protein e |
| 42 | 92.5 | 12.0 | 738 | 2 | AAW56163 | Aaw56163 New DNA s |
| 43 | 92.5 | 12.0 | 850 | 4 | ABR65764 | AbR65764 Drosophil |
| 44 | 92.5 | 12.0 | 1028 | 4 | ABR62708 | AbR62708 Drosophil |
| 45 | 92.5 | 12.0 | 1721 | 6 | ABU34624 | Abu34624 Protein e |

ALIGNMENTS

RESULT 1
AAB36355
ID AAB36355 standard; protein; 151 AA.
XX
AC AAB36355;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI: 2000-672631/65.
XX
DR N-PSDB; AAC64631.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 139; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively. (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 100.0%; Score 768; DB 3; Length 151;
 Best Local Similarity 100.0%; Pred. No. 1.7e-66;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNNATYDQLVTRVVTHEMAHAGG 120
 QY 121 NNAALVNOTASDSSVWVQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVWVQVGFNNATANQY 151

RESULT 2
 AAB36352
 ID AAB36352 standard; protein; 151 AA.
 AC AAB36352;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#7 amino acid sequence SEQ ID NO:24.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 XX Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.

PN WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 XX WPI; 2000-672631/65.
 DR N-PSDB; AAC64528.
 XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX

PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 91.1%; Score 700; DB 3; Length 151;
 Best Local Similarity 89.9%; Pred. No. 6.8e-60;
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
 QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNNAT------YDQLVTRVVT 112
 DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFNNATIDQWNAKNYDQLVTRVVT 120
 QY 113 HEMAHAGGNNALVNOTASDSSVWVQVGFNNATANQY 151
 DB 121 HEMAHA-----NQTASDSSVWVQVGFNNATANQY 151

RESULT 3
 AAR74625
 ID AAR74625 standard; protein; 151 AA.
 AC AAR74625;
 XX
 XX 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 XX Agfa sequence.

XX Salmonella; Agfa; vaccine.
 XX
 XX Salmonella.
 OS
 XX WO9425598-A2.
 XX
 XX 10-NOV-1994.
 PD
 XX 26-APR-1994; 94WO-IB000207.
 PF
 XX 26-APR-1993; 93US-00054452.
 PR
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX

```

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI; 1994-358275/44.
DR N-PSDB; AAO87467.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
XX strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7B; 95pp; English.
XX
XX The Salmonella AgfA protein and DNA are used in vaccine and genetic
XX immunization compositions, respectively, to elicit an immune response to
XX Salmonella in animals (e.g. food producing animals) and humans. (Updated
XX on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 151 AA;
Query Match 88.7%; Score 681; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 4.7e-58;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 4
AAB36341
ID AAB36341 standard; protein; 151 AA.
XX
XX AAB36341;
XX
XX 26-FEB-2001 (first entry)
XX
XX Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX
XX Salmonella enteritidis.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collinson SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX
XX N-PSDB; AAC64617.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant AgfA
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:

```

```

CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
CC assembly system of strains of Salmonella. Escherichia coli and
CC enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
Query Match 88.7%; Score 681; DB 3; Length 151;
Best Local Similarity 90.7%; Pred. No. 4.7e-58;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 5
AAW23570
ID AAW23570 standard; protein; 151 AA.
XX
XX AAW23570;
XX
XX 25-MAR-2003 (revised)
XX
XX 29-SEP-1997 (first entry)
XX
XX Salmonella enteritidis 27655-3b agfA.
XX
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.
XX
XX Salmonella enteritidis.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 123
XX FT /note= "Encoded by GCC"
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;

```

```

XX WPI; 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfa encoded by the full agfa gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteria. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridize to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
XX
XX Query Match 88.0%; Score 676; DB 2; Length 151;
XX Best Local Similarity 90.1%; Pred. No. 1.4e-57;
XX Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60
XX |||||
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60
XX |||||
XX QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFNNATYDQLVTRVVTHEMAHAGG 120
XX |||||
XX DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFNNATIDQWNAKNSDITVGYGG 120
XX |||||
XX QY 121 NNAALVNOTASDSSVWVRQVGFNNATANQY 151
XX |||||
XX DB 121 NNPALVNOTASDSSVWVRQVGFNNATANQY 151
XX |||||
XX
XX RESULT 6
XX AAB36347
XX ID AAB36347 standard; protein; 151 AA.
XX AC AAB36347;
XX
XX DT 26-FEB-2001 (first entry)
XX
XX DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.
XX
XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX
XX PN WO200060102-A2.
XX
XX PD 12-OCT-2000.
XX
XX PF 05-APR-2000; 2000WO-CA000356.
XX
XX PR 05-APR-1999; 99US-0127888P.
XX
XX PA (UYVI-) UNIV VICTORIA.
XX
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX DR WPI; 2000-672631/65.
XX DR N-PSDB; AAC64623.
XX
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa

```

protein useful for eliciting immune response in animal.
 Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/IAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 85.8%; Score 659; DB 3; Length 151;

Best Local Similarity 81.9%; Pred. No. 6.4e-56;

Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60

|||||

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSLTSLIYQYGSANAALALQ 60

|||||

QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFNNAT-----YDQ 105

|||||

DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFNNATIDQWNAKNSDITVGYDQ 120

|||||

QY 106 LVTRVVTHEMAHAGGNNALVNQTASDSVWVRQVGFNNATANQY 151

|||||

DB 121 LVTRVVTHEMAHAGGNNALVNQTASDSVWVRQVGFNNATANQY 151

|||||

RESULT 7

AAB36351

ID AAB36351 standard; protein; 151 AA.

XX AC AAB36351;

XX DT 26-FEB-2001 (first entry)

XX DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.

XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

XX vaccine; immune response; immunogen.

XX OS Salmonella enteritidis.

XX OS Escherichia coli.

XX OS Synthetic.

XX PN WO200060102-A2.

XX PD 12-OCT-2000.

XX PF 05-APR-2000; 2000WO-CA000356.

XX


```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64627.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 137; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating a replacement segment or segments of foreign amino
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ
Query Match 85.5%; Score 657; DB 3; Length 151;
Best Local Similarity 81.9%; Pred. No. 1e-55;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATYDQLVTRVVTHEMAHA-- 118
DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATYDQLVTRVVTHEMAHAQ 105
QY 119 -----CGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 106 WNAXNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 8
AAB36354
ID AAB36354 standard; protein; 151 AA.
XX
XX AAB36354;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#9 amino acid sequence SEQ ID NO:28.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64630.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ
Query Match 79.8%; Score 613; DB 3; Length 151;
Best Local Similarity 74.6%; Pred. No. 1.9e-51;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATYDQLVTRVVTHEMAHA-- 118
DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATYDQLVTRVVTHEMAHAFR 98
QY 119 -----CGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
DB 99 NNATIDQWAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
RESULT 9

```

AAB36346
 ID AAB36346 standard; protein; 151 AA.
 AC AAB36346;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64622.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;
 Query Match 78.5%; Score 603; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 1.7e-50;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 Qy 1 MKLLKVAFAAIVVSGSALACVFPQWGGGNGHNGSGPDSTLSIYQYGSANALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALACVFPQWGGGNGHNGSGPDSTLSIYQYGSANALALQ 60

Qy 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATYDQLVTRVWTHEAHAGG 120
 DB 61 SDARKSETTITQSGYNGADYVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 Qy 121 NNAALVNQTSADSSVMVROVGFNNATANQY 151
 DB 121 NNAALVNYDQLVTRVWTHEAHANNATANQY 151

RESULT 10

AAB36353
 ID AAB36353 standard; protein; 151 AA.
 XX
 AC AAB36353;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 DR N-PSDB; AAC64629.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (Sf17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in

```
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

  Query Match      78.1%; Score 600; DB 3; Length 151;
  Best Local Similarity 81.5%; Pred. No. 3.4e-50;
  Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
Dd 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALYDQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Dd 61 LVTRVVTHEMAHAGVGNADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Dd 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX

The present invention describes a recombinant agfa gene (I) where a
segment of the gene has been replaced by a segment of a foreign DNA
sequence which encodes a foreign epitope or antigen. Also described are:
(1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended
assembly system of strains of Salmonella, Escherichia coli and
Enterobacteriaceae for the production of fimbriae comprising recombinant
Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
directing recombination of a recombinant gene into the chromosome of the
homologous species; (3) directing recombination of a recombinant gene
back into the chromosome of the homologous species, replacing the native
copy of that gene; and (4) eliciting an immune response in an animal,
comprising separating an amino acid polymer comprising a recombinant Agfa
protein containing a replacement segment or segments of foreign amino
acid sequence or sequences grown on a Salmonella, E. coli or
```

```
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

  Query Match      78.0%; Score 599; DB 3; Length 151;
  Best Local Similarity 81.5%; Pred. No. 4.2e-50;
  Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
Dd 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDYDQLVTRVVTHEMAHALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Dd 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Dd 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
AAB36350
ID AAB36350 standard; protein; 151 AA.
XX
AC AAB36350;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX
OS Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
XX N-PSDB; AAC64626.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX

The present invention describes a recombinant agfa gene (I) where a
segment of the gene has been replaced by a segment of a foreign DNA
sequence which encodes a foreign epitope or antigen. Also described are:
```

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 77.7%; Score 597; DB 3; Length 151;
 Best Local Similarity 81.5%; Pred. No. 6.6e-50;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SDARKYDQLVTRVVTHEMAHAGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVQGYGG 120

QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151

RESULT 13
 AAB36348
 ID AAB36348 standard; protein; 151 AA.
 AC AAB36348;
 XX 26-FEB-2001 (first entry)
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 151 AA;

Query Match 73.7%; Score 566; DB 3; Length 151;
 Best Local Similarity 80.8%; Pred. No. 6.7e-47;
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVQGYGG 120

QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151

RESULT 14
 AAB36343
 ID AAB36343 standard; protein; 151 AA.
 AC AAB36343;
 XX 26-FEB-2001 (first entry)
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen.
 OS Escherichia coli.
 OS WO200060102-A2.
 XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UVI-1) UNIV VICTORIA.
XX
FI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR N-PSDB; AAC64619.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SeFl1/7AP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;
Query Match 67.3%; Score 517; DB 3; Length 151;
Best Local Similarity 69.5%; Pred. No. 3.8e-42;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 TDARNSDLTITQHGCGNGADVCGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVYVROVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
RESULT 15
ABR82651
ID ABR82651 standard; protein; 151 AA.
XX
AC ABR82651;
XX
DT 04-DEC-2003 (first entry)
XX
DE E. coli CsgA subunit 15 kDa protein.
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.
XX Escherichia coli.
XX WO2003064446-A2.
XX
PD 07-AUG-2003.
XX
XX 30-JAN-2003; 2003WO-EP000943.
XX
XX 31-JAN-2002; 2002GB-00002275.
XX
XX (HANS-) HANSA MEDICAL RES AB.
XX
PI Ejoerck L, Olsen A, Wikstroem M, Herwald H;
XX
XX WPI; 2003-646136/61.
DR N-PSDB; ACF36153.
XX
XX New isolated peptide capable of binding a mammalian plasma protein,
PT useful in the manufacture of a medicament for the prevention and/or
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella
PT or Shigella infections.
XX
XX Disclosure; Page 41-42; 42pp; English.
XX
XX The invention relates to an isolated peptide capable of binding a
CC mammalian plasma protein or of generating an immune response in a mammal
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or
CC antibody is useful for treating a bacterial infection in a human or
CC animal or in the manufacture of a medicament for the prophylactic
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella
CC or Shigella infection. The peptide that is immobilized on a solid support
CC is also useful as a reagent for determining the ability of a plasma
CC protein to bind to bacteria. The present sequence represents an E. coli
CC 15 kDa protein
XX
SQ Sequence 151 AA;
Query Match 66.7%; Score 512; DB 7; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-41;
Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 61 TDARNSDLTITQHGCGNGADVCGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNQTASDSSVYVROVGFNNATANQY 151
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151
Search completed: August 2, 2004, 14:48:29
Job time : 44.9 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds
(without alignments)
649.627 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 676 | 88.0 | 151 | 1 | US-08-233-788A-59 |
| 2 | 496 | 64.6 | 120 | 1 | US-08-233-788A-57 |
| 3 | 92.5 | 12.0 | 738 | 3 | US-08-864-038A-3 |
| 4 | 91.5 | 11.9 | 673 | 3 | US-08-196-387-8 |
| 5 | 91.5 | 11.9 | 673 | 4 | US-08-841-835-8 |
| 6 | 91.5 | 11.9 | 949 | 3 | US-08-196-387-10 |
| 7 | 91.5 | 11.9 | 949 | 4 | US-08-841-835-10 |
| 8 | 91.5 | 11.9 | 1327 | 3 | US-08-196-387-2 |
| 9 | 91.5 | 11.9 | 1327 | 4 | US-08-841-835-2 |
| 10 | 91.5 | 11.9 | 1327 | 4 | US-08-972-115A-8 |
| 11 | 86.5 | 11.3 | 745 | 4 | US-08-336-115C-6 |
| 12 | 84 | 10.9 | 2736 | 4 | US-08-252-991A-30227 |
| 13 | 83 | 10.8 | 1216 | 4 | US-08-134-000C-5130 |
| 14 | 81.5 | 10.6 | 589 | 4 | US-08-489-039A-7849 |
| 15 | 80 | 10.4 | 867 | 4 | US-08-540-236-2676 |
| 16 | 80 | 10.4 | 878 | 4 | US-08-540-236-3401 |
| 17 | 78.5 | 10.2 | 309 | 4 | US-08-252-991A-22266 |
| 18 | 78.5 | 10.2 | 873 | 4 | US-08-336-447A-13 |
| 19 | 78.5 | 10.2 | 892 | 4 | US-08-336-447A-5 |
| 20 | 78.5 | 10.2 | 1415 | 4 | US-08-252-991A-26438 |
| 21 | 77.5 | 10.1 | 1160 | 3 | US-08-808-599A-24 |
| 22 | 77.5 | 10.1 | 1207 | 4 | US-08-489-039A-11518 |
| 23 | 77 | 10.0 | 415 | 4 | US-08-025-769B-280 |
| 24 | 77 | 10.0 | 975 | 4 | US-08-328-352-4764 |
| 25 | 77 | 10.0 | 1003 | 1 | US-08-571-758-4 |
| 26 | 77 | 10.0 | 1003 | 1 | US-08-909-984A-4 |
| 27 | 77 | 10.0 | 1003 | 1 | US-08-909-983-4 |

| | | | | | | |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 77 | 10.0 | 1034 | 4 | US-09-252-991A-26658 | Sequence 26658, A |
| 29 | 77 | 10.0 | 1890 | 4 | US-09-595-684B-39 | Sequence 39, Appl |
| 30 | 77 | 10.0 | 1864 | 2 | US-08-804-227C-3 | Sequence 3, Appl |
| 31 | 77 | 10.0 | 2315 | 4 | US-09-543-681A-5434 | Sequence 5434, Ap |
| 32 | 76.5 | 10.0 | 624 | 4 | US-09-336-447A-7 | Sequence 7, Appl |
| 33 | 76.5 | 10.0 | 941 | 4 | US-09-336-447A-9 | Sequence 9, Appl |
| 34 | 76.5 | 10.0 | 1398 | 1 | US-08-750-532-9 | Sequence 9, Appl |
| 35 | 76.5 | 10.0 | 1398 | 3 | US-08-894-818B-8 | Sequence 8, Appl |
| 36 | 76.5 | 10.0 | 1398 | 4 | US-09-445-472-6 | Sequence 6, Appl |
| 37 | 76.5 | 10.0 | 1612 | 1 | US-08-169-927-2 | Sequence 2, Appl |
| 38 | 76.5 | 10.0 | 2123 | 3 | US-08-568-685A-10 | Sequence 10, Appl |
| 39 | 76 | 9.9 | 231 | 4 | US-09-540-236-3827 | Sequence 3827, Ap |
| 40 | 76 | 9.9 | 870 | 4 | US-09-177-650-91 | Sequence 91, Appl |
| 41 | 75.5 | 9.8 | 238 | 4 | US-09-495-880A-42 | Sequence 42, Appl |
| 42 | 75.5 | 9.8 | 282 | 2 | US-08-860-174A-10 | Sequence 10, Appl |
| 43 | 75.5 | 9.8 | 365 | 4 | US-09-489-847-322 | Sequence 322, App |
| 44 | 75.5 | 9.8 | 394 | 4 | US-09-252-991A-28148 | Sequence 28148, A |
| 45 | 75.5 | 9.8 | 432 | 4 | US-09-403-089A-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-233-788A-59
; Sequence 59, Application US/08233788A
; Patent No.: 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDABERRY
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-59

Query Match 88.0%; Score 676; DB 1; Length 151;
Best Local Similarity 90.1%; Pred. No. 2.7e-61;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAIVVPGVGGGNNHNGSGDPDSTLTIYQYSANALALQ 60
|||||

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGGNSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNQTSADSSVMVQVGFNNATANQY 151
Db 121 NNPALVNQTSADSSVMVQVGFNNATANQY 151
RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Key, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouchier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57
Query Match 64.6%; Score 496; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 4e-43;
Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;
QY 22 VVPQGGGNNHGGGNSGSPDSTLSIYQYGSANAALALQSPARKSETTITQSGYNGADY 81
Db 1 VVPQGGGNNHGGGNSGSPDSTLSIYQYGSANAALALQSPARKSETTITQSGYNGADY 60
QY 82 GQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNNALVNQTSADS 133
Db 61 GQGDNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTSADS 112
RESULT 3
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF

; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isehinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 986-2340
; TELEFAX: (212) 953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
US-08-864-038A-3
Query Match 12.0%; Score 92.5; DB 3; Length 738;
Best Local Similarity 27.7%; Pred. No. 0.46; Indels 39; Gaps 6;
Matches 43; Conservative 12; Mismatches 61; Indels 39; Gaps 6;
QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNNHGGGNSGSPDSTLSIYQYGSANAALALQSD 62
Db 419 LLKSSASASASASASASAG-----GGGGGGGGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYNGADYQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNN 122
Db 461 -----AALAAAGAGGGGGLGGGGGAGALAAALAAAG-AGGGGGGGL-----GGGLGGLGGGS 508
QY 123 AALVNQTSADSS-----VMVQVGFNNATA 148
Db 509 AAAAAAASASASASASASAG-----GGGGGGGGGGGGG-----GGGGGAGALA----- 543
RESULT 4
US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF


```
;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 3; Length 673;
Best Local Similarity 29.4%; Pred. No. 0.51;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGNGHNGSGDPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSTAPLPGAGPGTGVPVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVFG 143
Db 207 VNAAK--DMAGRKSSPLHFAAGFG 227

RESULT 5
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.

;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-196-387-8

Query Match 11.9%; Score 91.5; DB 3; Length 673;
Best Local Similarity 29.4%; Pred. No. 0.51;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGNGHNGSGDPDSTLSIYQVGSANAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSTAPLPGAGPGTGVPVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVFG 143
Db 207 VNAAK--DMAGRKSSPLHFAAGFG 227

RESULT 6
US-09-196-387-10
; Sequence 10, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
```

```
;
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-196-387-10

Query Match 11.9%; Score 91.5; DB 3; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVQGGADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGGAAGPGTGPVAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAK--DMAGRKSSPLHPAAGFG 227

RESULT 7
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
;
; MOLECULE TYPE: protein
US-09-841-835-10

Query Match 11.9%; Score 91.5; DB 4; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGDPSTLSIYQYGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVQGGADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGGAAGPGTGPVAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAK--DMAGRKSSPLHPAAGFG 227

RESULT 8
US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 627613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-196-387-2

Query Match 11.9%; Score 91.5; DB 3; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNSGDPSTLSIYQYGSANAALALQSDAR 64
```

Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNSSSSPTSS--SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADYQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGPGTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGRKSSPLHFAAGFG 227

RESULT 9
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/841.835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-841-835-2

Query Match
Best Local Similarity 11.9%; Score 91.5; DB 4; Length 1327;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;
QY 6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSGPDSTLSIYQVGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNSSSSPTSS--SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADYQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGPGTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGRKSSPLHFAAGFG 227

RESULT 10
US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Patent No. 6599728
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gregg, Morin B.
; APPLICANT: Walter, Funk D.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match
Best Local Similarity 11.9%; Score 91.5; DB 4; Length 1327;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;
QY 6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSGPDSTLSIYQVGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNSSSSPTSS--SSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADYQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGPGTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
Db 207 VNAX--DMAGRKSSPLHFAAGFG 227

RESULT 11
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; TITLE OF INVENTION: Methods Against Helicobacter Infection
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE: VARIANT
; NAME/KEY: 721
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

```
Query Match      11.3%; Score 66.5; DB 4; Length 745;
Best Local Similarity 20.0%; Pred. No. 1.9;
Matches 34; Conservative 27; Mismatches 72; Indels 37; Gaps 5;

QY 8 AFAATVSGSALAGVVPW-----GGGNGHNGGN----- 37
DB 102 AYQAVFLAINAAVL--WNTIGYAVMCGNGGTSGPVSIFNDQPCQDSTQITCNKFE 158

QY 38 SSGPDSTLSIYQGSANAALQSDARKSETTITQSGVGCADYQCGADNSTIELTQNGF 97
DB 159 STGPKGMSIDEFKXLEAYQIIQALKNQSGFFELG--GNQTKV---SVNYYNECROTAD 214

QY 98 RNNATYDQLVTRVTHEMAHAGGNAALVNQTSDDSSVMVRQVFGNNAT 147
DB 215 INGGVYQCKAKNGSSSSNGGSSSTQTTATTQDGVITITTYNNKAT 264

RESULT 12
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match      10.9%; Score 84; DB 4; Length 2736;
Best Local Similarity 29.7%; Pred. No. 19;
Matches 43; Conservative 17; Mismatches 49; Indels 36; Gaps 8;

QY 6 VAAFAIVVSGSAL-AGVV--PQWGGGNGHNGGNSGPDST-----LSIYQVGSANA 55
DB 149 MSAGTALLVGAAGAGVAINSSGGG---GGGSSVPPDTPPKAASGLKIAPDGSIS 205

QY 56 ALALQSDARKSETTITQSGVGCADYV-CQGADNSTIELTQNG-----FRNNATYDQLVTR 109
DB 206 GQA-----EAGASGVIDTNGDGKPLDVIADANGFTAPLNPLTNGQVTV 252

QY 110 VVTHEMAHAGGNAALVNQTSDDS 134
DB 253 VVT-----DPAGNASPPAQVTAFTDT 273

RESULT 13
US-09-134-000C-5130
; Sequence 5130, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5130
; LENGTH: 1216
```

```
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5130

Query Match      10.8%; Score 83; DB 4; Length 1216;
Best Local Similarity 24.4%; Pred. No. 8.2;
Matches 49; Conservative 19; Mismatches 55; Indels 78; Gaps 10;

QY 6 VAAFAIVV--SGSALAGVVPWGG--GSHNGGNSGPDST-----LSIY- 48
DB 181 VTGNSLVNDSSGGRIAS---YGGYGNTATNTANTVGNSTKVAITNAATGFOLSTYY 237

QY 49 ---QYGSANAALQSDARKSETTITQSGVGCADYVQ-----GADNSTIELTQ 94
DB 238 GGVOYGNIG-----GKVTNDISGYGRWYTAGORFIGGSSRGDIGNRATDGT 286

QY 95 N-----GFRNNATYDQLVTRVTHEMAHAG-----GNNAAALVNQTA 130
DB 287 NLNTQLYSAGRADFEFGNQYSGTIIGDITNVVTAGTNSAGGINDFNGGAGNNVSKFNK- 345

QY 131 SDSSVMVRQVFGNNATANCY 151
DB 346 -----QIGASNEATYDAY 358

RESULT 14
US-09-489-039A-7849
; Sequence 7849, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7849
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7849

Query Match      10.6%; Score 81.5; DB 4; Length 589;
Best Local Similarity 24.5%; Pred. No. 4.5;
Matches 35; Conservative 21; Mismatches 56; Indels 31; Gaps 7;

QY 29 GGNHNGGNSGPDSTLSIYQGS-----ANAALQSDARKSETTITQSGYNGA- 79
DB 259 GGLDRNGANANGQDITFGIYAFDITLTLTIERIEINGGLRLDNYHTKYDSATACGSGRGAI 318

QY 80 --DVQ--GADNSTIELTONG-----FRNNATYDQLVTRVTHEMAHAGN---NAALVNQ 128
DB 319 ACPFGQSTGSPVTVDTAKSGLNVNWKAGALYR-----LTBQ-----GNVYVNTAISQQ 367

QY 129 TASDSSVMVRQVFGNNATANCY 151
DB 368 PPGGSSFALAAASGSGNSANRTDF 390

RESULT 15
US-09-540-236-2676
; Sequence 2676, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT FILING DATE: 2000-04-04
```

; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2676
; LENGTH: 867
; TYPE: PRT
; ORGANISM: M.cattarrhalis
US-09-540-236-2676

Query Match 10.4%; Score 80; DB 4; Length 867;
Best Local Similarity 26.8%; Pred. No. 11;
Matches 49; Conservative 23; Mismatches 65; Indels 46; Gaps 10;
QY 4 LKVAFAAIVVSSGALAG--VVPQWGG-----GGNHN-----GGG---NSSGPD 43
DB 33 LGMATTASAOVQGAIVGGNRNVNQATGKYSTVSGGDYNEANGEISTIGGGFFNKASGESS 92
QY 44 TLSIVQYGSNAALALQSDARKSETTITQSGY----GNGADVGGADNST--IELTQNGF 97
DB 93 TIA-----GGRN-----NQATKENSTVGGGKFNQAKGRNSTVAGGYNNEATGIDSTIAGG 142
QY 98 R-NNATYDQLVTRVTHEMAHAGNNAALVNTASDSSVMV-----RQVGFNNAT 147
DB 143 RGNQATGAGSFAAGVGNQATGAGSFAAGVGNQANADNAVAVGNKNSITGKDSVAIGSNNT 202
QY 148 ANQ 150
DB 203 VAQ 205

Search completed: August 2, 2004, 14:58:37
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds
(without alignments)

1287.123 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSVMVRQVFGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 514 | 66.9 | 151 | 12 | US-09-741-873B-4 |
| 2 | 514 | 66.9 | 151 | 12 | US-09-741-873B-4 |
| 3 | 436 | 56.8 | 131 | 12 | US-09-741-873B-2 |
| 4 | 436 | 56.8 | 131 | 12 | US-09-741-873B-2 |
| 5 | 103 | 13.4 | 445 | 15 | US-10-369-493-20638 |
| 6 | 94.5 | 12.3 | 537 | 16 | US-10-437-963-141342 |
| 7 | 94 | 12.2 | 354 | 10 | US-09-820-843A-21 |
| 8 | 94 | 12.2 | 498 | 12 | US-10-424-599-196154 |
| 9 | 93.5 | 12.2 | 2204 | 12 | US-10-282-122A-64364 |
| 10 | 93 | 12.1 | 263 | 12 | US-10-425-114-49960 |
| 11 | 92.5 | 12.0 | 1721 | 12 | US-10-282-122A-62548 |
| 12 | 91.5 | 11.9 | 673 | 9 | US-09-841-835-8 |
| 13 | 91.5 | 11.9 | 949 | 9 | US-09-841-835-10 |
| 14 | 91.5 | 11.9 | 1327 | 9 | US-09-841-835-2 |
| 15 | 91.5 | 11.9 | 1327 | 10 | US-09-972-115A-8 |

```
Sequence 4, Appli
Sequence 53269, A
Sequence 173860,
Sequence 108981, A
Sequence 48048, A
Sequence 147748,
Sequence 2, Appli
Sequence 49757, A
Sequence 105413,
Sequence 56041, A
Sequence 275468, A
Sequence 57763, A
Sequence 146, App
Sequence 67750, A
Sequence 6, Appli
Sequence 58683, A
Sequence 68152, A
Sequence 122263, A
Sequence 47486, A
Sequence 1074, App
Sequence 172276,
Sequence 590, App
Sequence 26, Appl
Sequence 26, Appl
Sequence 749, App
Sequence 24, Appl
Sequence 164573, A
Sequence 16, Appl
Sequence 33, Appl
Sequence 115033,
US-09-741-873B-4
US-10-199-937-4
US-10-282-122A-53269
US-10-424-599-173860
US-10-437-963-108981
US-10-282-122A-48048
US-10-437-963-147748
US-10-311-406-2
US-10-282-122A-49757
US-10-437-963-105413
US-10-425-114-56041
US-10-424-599-275468
US-10-425-114-57763
US-09-793-306-146
US-10-425-114-67750
US-08-834-666A-6
US-10-282-122A-58683
US-10-425-114-68152
US-10-437-963-122263
US-10-282-122A-47486
US-10-412-699B-1074
US-10-374-780A-590
US-10-437-963-172276
US-10-437-963-26
US-10-021-811-26
US-10-238-075-749
US-09-905-176-24
US-10-282-122A-64573
US-09-996-134-16
US-10-164-966-33
US-10-437-963-115033
```

ALIGNMENTS

RESULT 1

```
US-09-741-873B-4
; Sequence 4; Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4
```

Query Match 66.9%; Score 514; DB 12; Length 151;

Best Local Similarity 68.9%; Pred. No. 1.4e-43;

Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALGVVQWGGGNGGNSGPDSTLSIYQVGSANAALQ 60

| | | | |
|----|-----|--|-----|
| Db | 1 | MLLLKVAATAATVFGSAVAGVWPQYGGGNGHGGGNGSGFSELNIYQYGGGNSALAQ | 60 |
| Qy | 61 | SDARKSETITQSGYGCNGADYCGQAGDNSTIELTQGFRRNATYDQLVTRVVTHEMAHAGG | 120 |
| Db | 61 | TDARNSDLTITGGGNGADYCGQSDSSIDLTQGFNGSATLDQWNGKNSMTVKVQFG | 120 |
| Qy | 121 | NNAALVNQTSADSSVNVQVFGFNATANQY | 151 |
| Db | 121 | GRGAAVDQTSANSSVNVTVQVFGFNATAHQY | 151 |

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsner, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B

```

/ CURRENT FILING DATE: 2003-04-04
/ PRIOR APPLICATION NUMBER: SE 8801723-1
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: US 08/978,878
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: US 07/347,189
/ PRIOR FILING DATE: 1989-05-04
/ PRIOR APPLICATION NUMBER: US 07/789,437
/ PRIOR FILING DATE: 1991-11-06
/ PRIOR APPLICATION NUMBER: US 07/970,846
/ PRIOR FILING DATE: 1992-11-03
/ PRIOR APPLICATION NUMBER: US 08/187,865
/ PRIOR FILING DATE: 1994-01-28
/ PRIOR APPLICATION NUMBER: US 08/318,519
/ PRIOR FILING DATE: 1994-10-05
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 4
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Escherichia coli
/ US-09-741-873B-4

```

```

Query Match      66.9%; Score 514; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. 1.4e-43;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1  MKLLKVAFAAIVVSGSALAGVTPQWGGGHNHGGNSGPGSTLSIYYGGSNAAALQ 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1  MKLLKVAFAAIAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYYGGNSALQ 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      61  SPARKSETITIQSGYNGADVGGCADNSTIELTQNGFRNNATYDOLVTRVWTHENAHAGG 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61  TDARNSDLITIQHGGGNGADVGGSDSSIDLTRQFGNSAILDQWNGKNSMTVKQFGG 120
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      121  NNAALVNOTASDSSVMYRVQVFGNNATAOY 151
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      121  CNGAAVDOTASNSVMYTVQVFGNNATAHOY 151
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olser, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012869-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication NO. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2
Query Match          56.8%      Score 436;   DB 12;   Length 131;

```



```
Best Local Similarity 65.6%; Pred. No. 7.7e-36;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

Qy 21 GVVPQGGGNGHNGSGSPDSTLSIYQVGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQGGGNGHNGSGSPDSTLSIYQVGSANAALALQSDARKSETTITQSGYNGAD 60

Qy 81 VQGGADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAGGNNALVNQTSDDSSVMVQV 140
Db 61 VQGGSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGGNGAAVDQTSNSSVNVTVQ 120

Qy 141 GFGNNATANQY 151
Db 121 GFGNNATAHOY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; NAME/KEY: unsure
; FEATURE:
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.4%; Score 103; DB 15; Length 445;
Best Local Similarity 26.3%; Pred. No. 0.079;
Matches 44; Conservative 27; Mismatches 62; Indels 34; Gaps 8;

Qy 7 AAFAA-----IVVSGSALAGVWPQWGGG-----NHNGG-----GNSGPDSTLSIYQY 50
Db 19 AAFRAADSNVYLNQTDGNDQANITQSGNGSVGAFNGSGFLQENGTLGA-NLLTVKQS 77

Qy 51 GSANAALALQSDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNA-----TY 103
Db 78 GNSNSV-----GRDIQGGQSGAGNSAIFQEGTQSDVELQQTGTGNSGAVFSGWNTN 129

Qy 104 DQLVTRVVTTHMAHAGGNNALVNQTSDDSSVMVQVQFGNNATANQ 150
Db 130 DPGVFNKTIQD-SSNSGKSVI-QDGKNVFSIKQGTNGTNSVQ 174

RESULT 6
US-10-437-963-141342
; Sequence 141342, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
```

```
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 141342
; LENGTH: 537
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_42455C.1.pap
US-10-437-963-141342

Query Match 12.3%; Score 94.5; DB 16; Length 537;
Best Local Similarity 27.1%; Pred. No. 0.71;
Matches 45; Conservative 17; Mismatches 55; Indels 49; Gaps 9;

Qy 3 LLKVAAPAAIIVVSGSALAGVWPQWGGG-----GNHNGG-----NSS 39
Db 347 LLNIGFF-----SGS--NGGYDQFNGGAGNGGGSIVTSSGLAGNHGGGGFPLYNSS 399

Qy 40 QPDSTLSIYQVGSANAALALQSDARKSETTITQSGYNGADYVQGGADNSTIELTQNGF 97
Db 400 EPAGTLP-----QMSATALLQAAQMGATTSSYNAGGAGGASSLLRGASSHGISVGECPA 454

Qy 98 RNNATYDQLVTRVVTTHMAHAGGNNALVNQTSDDSSVMVQVQVGF 143
Db 455 NERSYQNL-----IMGSVA-SGGGAGPAGFSGAS-----GEG 488

RESULT 7
US-09-820-843A-21
; Sequence 21, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTE
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 354
; TYPE: PRT
; ORGANISM: M. tuberculosis
; NAME/KEY: misc.feature
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: gi|1781260
US-09-820-843A-21

Query Match 12.2%; Score 94; DB 10; Length 354;
Best Local Similarity 25.0%; Pred. No. 0.47;
Matches 35; Conservative 16; Mismatches 71; Indels 18; Gaps 5;

Qy 26 WGGGNGHNGGNSGPDSS---TLSIYQVGSANAALALQSDARKSETTI-----TQSGY 75
Db 148 WNGGSTNTGLANAGAGNTGFFDAGNYNFGSLNAGNINSGNSGDNGLNAGDVNSGV 207

Qy 76 GNGADYVQGGADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAG-----GNNALVNQ 129
Db 208 GNAGDVNTGLNSG-NINTGGF-NPGLTNTGFFSANTQAGPNSGFFNAGTGNSGFGHNDP 265

Qy 130 ASDSSVMVQVQFGNNATAN 149
Db 266 AGSGNSGIONSGFGNSGYVN 285
```

```

RESULT 8
US-10-424-599-196154
; Sequence 196154, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196154
; LENGTH: 498
; TYPE: PRF
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19152C.i.pep
US-10-424-599-196154

Query Match          12.2%; Score 94; DB 12; Length 498;
Best Local Similarity 28.1%; Pred. No. 0.72;
Matches      34; Conservative    24; Mismatches   41; Indels     22; Gaps       6;

QY      30  GNHGGGNSGGPD-STLSIYQYGSAUALALQSADARKSETTITQSGYGNG-----A 79
         |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      355  GGSNDSNNASQSDTSSEQNKGSSNDNS--NDANQNGSNGENANDNCNASHEAQNNA 412
         |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      80  DVCGGAGDNSTIEL----TQGFENNATYDOLVTRVVTHMAHAGG-NNAALVNOTASDSS 134
         |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      413  DACGGNNEATIVESKASENEGGAQTIVESQ-----KEEESAHSGGDSNLNDQGSSDHS 467
         |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      135  V 135
         |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      468  I 468
         |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 9
US-10-282-122A-64364
; Sequence 64364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malore, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64364
; LENGTH: 2204
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64364

Query Match          12.2%; Score 93.5; DB 12; Length 2204;
Best Local Similarity 26.1%; Pred. No. 5.1;
Matches 43; Conservative 18; Mismatches 63; Indels 41; Gaps 6;

Qy      10 AATVSG-----SALGV-----VQVGGGNHNGGNSGDPDSTLSIYQVCSA 53
Db      840 AATVGGPTTSIGTASAGISITPIIDIPATSGFGN-----STTSPSGGFNNGSGA 894

Qy      54 NAALALQSDARKSETITQSYGNGADVGQADN-----STIELTQNGFRNNAT 102
Db      895 SGFLNVVAGASGI-----SGVLNVGALGSGVTNVGHTVSGFYNASALDLVTPAFASGLM 948

Qy      103 YDCLVTRVVTHEMAHGAGNNALVNCQTASDSVM---VRQVGCN 144
Db      949 RDGMGTMTLNLGLANLGSNNAGFGNTGTFDVGVANLGNYNIGFN 993

RESULT 10
US-10-425-114-49960
; Sequence 4960, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jirgdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21/53913/B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49960
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: C clone ID: 700071884_FLI.pep
US-10-425-114-49960

```

[illegible]

Db 188 SGA 190

RESULT 11

US-10-282-122A-62548
; Sequence 62548, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62548
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62548

Query Match 12.0%; Score 92.5; DB 12; Length 1721;
Best Local Similarity 29.5%; Pred. No. 4.8;
Matches 39; Conservative 18; Mismatches 58; Indels 17; Gaps 6;

Qy 23 VPQWGGGNGHNGGNS-----SGPDSTLSIQY-GSANAALALQSDAKSETTITQSGYGN 77
Db 921 IPATPGFGNSTGGUSSGFFNSGAGSGFGNFGGAAGFNNLVS-----TTSGMSGFLN 974

Qy 78 GADYQGQADNSTIELTQGFRRNATYDQLVTRVVTTHMAHAGGNNAALVNQTASDSSVMV 137
Db 975 VGALGSGVAN--VGNITISGIYVGTSD-LSTPAVNSGLANIGTNIAGLLRDGAGTAAI-- 1029

Qy 138 RQVFGGNNTAN 149
Db 1030 -NLGLAHGNLN 1040

RESULT 12

US-09-841-835-8
; Sequence 8, Application US/09841835

Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 9; Length 673;
Best Local Similarity 29.4%; Pred. No. 1.9;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

Qy 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIQYGSANAALALQSDAR 64
Db 99 VAAAFVVPVAVSTSSAAGVAPNPAGSGNSPSSSSPTSS-SSSSPSFGSLSAESPEAA 157

Qy 65 KSETTIT---QSGYNGADYQGQADNSTIELTQGFRRNATYDQLVTRVVTTHMAHAGG 120
Db 158 GVSSTAPLPGGAAGPGTGVPVSGALRELLACRNGD-----VSRV--KRLVDAAN 206

Qy 121 NNAALVNQTASDSSVMVQVGF 143
Db 207 VNAAK--DMAGRRKSPHPAAGFG 227

RESULT 13

US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack

```

; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-841-835-10

Query Match 11.9%; Score 91.5; DB 9; Length 949;
Best Local Similarity 29.4%; Pred. No. 2.9;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAFPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

QY 65 KSETTIT-----QSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAAGGTGTPAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206

QY 121 NNAALVNQTASDSSVMVRQVGFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 14
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. US20020076795A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-841-835-2

Query Match 11.9%; Score 91.5; DB 9; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAFPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

QY 65 KSETTIT-----QSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAAGGTGTPAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206

QY 121 NNAALVNQTASDSSVMVRQVGFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 15
US-09-972-115A-8
; Sequence 8, Application US/09972115A
; Publication No. US20030032769A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Walter, Mark B.
; APPLICANT: Mieczyslaw, Piatyszek A.
; TITLE OF INVENTION: A Second Mammalian Telomerase
; FILE REFERENCE: 080/003C
; CURRENT APPLICATION NUMBER: US/09/972,115A
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/128,577
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: US 60/129,123
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-972-115A-8

Query Match 11.9%; Score 91.5; DB 10; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAFPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157

QY 65 KSETTIT-----QSGYNGADVGQGDNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120

```

Db 158 GVSSTAPLPGAGPGTGVPAVSGALRELLLEACKNGD-----VSRV--KRLVDAAN 206
Qy 121 NNAALVNQTASDSSVMVRQVGFG 143
Db 207 VNAK--DNAGRKSSFLHFAAGFG 227

Search completed: August 2, 2004, 15:36:14
Job time : 37.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds
(without alignments)
877.809 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main.*
1: /cgn2_6/ptodata/2/paa/PCTUS COMB.pcp.*
2: /cgn2_6/ptodata/2/paa/US06 COMB.pcp.*
3: /cgn2_6/ptodata/2/paa/US07 COMB.pcp.*
4: /cgn2_6/ptodata/2/paa/US08 COMB.pcp.*
5: /cgn2_6/ptodata/2/paa/US081 COMB.pcp.*
6: /cgn2_6/ptodata/2/paa/US082 COMB.pcp.*
7: /cgn2_6/ptodata/2/paa/US083 COMB.pcp.*
8: /cgn2_6/ptodata/2/paa/US084 COMB.pcp.*
9: /cgn2_6/ptodata/2/paa/US085 COMB.pcp.*
10: /cgn2_6/ptodata/2/paa/US086 COMB.pcp.*
11: /cgn2_6/ptodata/2/paa/US087 COMB.pcp.*
12: /cgn2_6/ptodata/2/paa/US088 COMB.pcp.*
13: /cgn2_6/ptodata/2/paa/US089 COMB.pcp.*
14: /cgn2_6/ptodata/2/paa/US090 COMB.pcp.*
15: /cgn2_6/ptodata/2/paa/US091 COMB.pcp.*
16: /cgn2_6/ptodata/2/paa/US092 COMB.pcp.*
17: /cgn2_6/ptodata/2/paa/US093 COMB.pcp.*
18: /cgn2_6/ptodata/2/paa/US094 COMB.pcp.*
19: /cgn2_6/ptodata/2/paa/US095 COMB.pcp.*
20: /cgn2_6/ptodata/2/paa/US096 COMB.pcp.*
21: /cgn2_6/ptodata/2/paa/US097A COMB.pcp.*
22: /cgn2_6/ptodata/2/paa/US097B COMB.pcp.*
23: /cgn2_6/ptodata/2/paa/US098 COMB.pcp.*
24: /cgn2_6/ptodata/2/paa/US099A COMB.pcp.*
25: /cgn2_6/ptodata/2/paa/US099B COMB.pcp.*
26: /cgn2_6/ptodata/2/paa/US100 COMB.pcp.*
27: /cgn2_6/ptodata/2/paa/US101 COMB.pcp.*
28: /cgn2_6/ptodata/2/paa/US102 COMB.pcp.*
29: /cgn2_6/ptodata/2/paa/US103 COMB.pcp.*
30: /cgn2_6/ptodata/2/paa/US104 COMB.pcp.*
31: /cgn2_6/ptodata/2/paa/US106 COMB.pcp.*
32: /cgn2_6/ptodata/2/paa/US107 COMB.pcp.*
33: /cgn2_6/ptodata/2/paa/US60 COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
|------------|-------|-------|--------|-------|-------------|

| 1 | 768 | 100.0 | 151 | 19 | US-09-543-407-30 | Sequence 30, Appl |
|----|-------|-------|------|----|---------------------|-------------------|
| 2 | 700 | 91.1 | 151 | 19 | US-09-543-407-24 | Sequence 24, Appl |
| 3 | 681 | 88.7 | 151 | 19 | US-09-543-407-5 | Sequence 5, Appl |
| 4 | 676 | 88.0 | 151 | 6 | US-08-233-642A-57 | Sequence 57, Appl |
| 5 | 659 | 85.8 | 151 | 19 | US-09-543-407-14 | Sequence 14, Appl |
| 6 | 657 | 85.5 | 151 | 19 | US-09-543-407-22 | Sequence 22, Appl |
| 7 | 613 | 78.8 | 151 | 19 | US-09-543-407-28 | Sequence 28, Appl |
| 8 | 603 | 78.5 | 151 | 19 | US-09-543-407-12 | Sequence 12, Appl |
| 9 | 600 | 78.1 | 151 | 19 | US-09-543-407-26 | Sequence 26, Appl |
| 10 | 599 | 78.0 | 151 | 19 | US-09-543-407-18 | Sequence 18, Appl |
| 11 | 597 | 77.7 | 151 | 19 | US-09-543-407-20 | Sequence 20, Appl |
| 12 | 594 | 77.3 | 151 | 19 | US-09-543-407-31 | Sequence 31, Appl |
| 13 | 566 | 73.7 | 151 | 19 | US-09-543-407-16 | Sequence 16, Appl |
| 14 | 517 | 67.3 | 151 | 19 | US-09-543-407-7 | Sequence 7, Appl |
| 15 | 514 | 66.9 | 151 | 13 | US-08-978-878-4 | Sequence 4, Appl |
| 16 | 514 | 66.9 | 151 | 21 | US-09-741-873B-4 | Sequence 4, Appl |
| 17 | 512 | 66.7 | 151 | 33 | US-60-352-946-2 | Sequence 2, Appl |
| 18 | 512 | 66.7 | 151 | 33 | US-60-444-371-2 | Sequence 2, Appl |
| 19 | 496 | 64.6 | 120 | 6 | US-08-233-642A-55 | Sequence 55, Appl |
| 20 | 475 | 61.8 | 158 | 16 | US-09-252-691-5834 | Sequence 5834, Ap |
| 21 | 475 | 61.8 | 158 | 16 | US-09-252-691C-5834 | Sequence 5834, Ap |
| 22 | 475 | 61.8 | 158 | 30 | US-10-417-886-5834 | Sequence 34, Appl |
| 23 | 462 | 60.2 | 109 | 19 | US-09-543-407-34 | Sequence 2, Appl |
| 24 | 436 | 56.8 | 131 | 13 | US-08-978-878-2 | Sequence 35, Appl |
| 25 | 436 | 56.8 | 131 | 21 | US-08-741-873B-2 | Sequence 37, Appl |
| 26 | 332 | 43.2 | 109 | 19 | US-09-543-407-35 | Sequence 39, Appl |
| 27 | 310 | 40.4 | 68 | 19 | US-09-543-407-37 | Sequence 32, Appl |
| 28 | 237 | 30.9 | 48 | 19 | US-09-543-407-39 | Sequence 5833, Ap |
| 29 | 217.5 | 28.3 | 70 | 19 | US-09-543-407-32 | Sequence 5833, Ap |
| 30 | 114.5 | 14.9 | 186 | 16 | US-09-252-691-5833 | Sequence 5833, Ap |
| 31 | 114.5 | 14.9 | 186 | 16 | US-09-252-691C-5833 | Sequence 5833, Ap |
| 32 | 114.5 | 14.9 | 186 | 30 | US-10-417-886-5833 | Sequence 8, Appl |
| 33 | 104.5 | 13.6 | 151 | 19 | US-09-543-407-8 | Sequence 20638, A |
| 34 | 103 | 13.4 | 445 | 29 | US-10-369-493-20638 | Sequence 20638, A |
| 35 | 103 | 13.4 | 445 | 33 | US-60-360-039-20638 | Sequence 21, Appl |
| 36 | 101.5 | 13.2 | 520 | 1 | PCT-US02-18256-21 | Sequence 6, Appl |
| 37 | 99.5 | 13.0 | 151 | 19 | US-09-543-407-6 | Sequence 8854, Ap |
| 38 | 98 | 12.8 | 145 | 21 | US-09-739-449-8854 | Sequence 8854, Ap |
| 39 | 98 | 12.8 | 145 | 23 | US-09-803-110-8854 | Sequence 22325, A |
| 40 | 98 | 12.8 | 375 | 21 | US-09-733-089-22325 | Sequence 22325, A |
| 41 | 98 | 12.8 | 375 | 23 | US-09-816-660-22325 | Sequence 23333, A |
| 42 | 98 | 12.8 | 563 | 21 | US-09-733-089-23333 | Sequence 23333, A |
| 43 | 98 | 12.8 | 563 | 23 | US-09-816-660-23333 | Sequence 9129, Ap |
| 44 | 96.5 | 12.6 | 313 | 21 | US-09-708-427-9129 | Sequence 359, App |
| 45 | 96.5 | 12.6 | 1249 | 30 | US-10-455-719-358 | |

ALIGNMENTS

RESULT 1
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-30

| | | | | | | | |
|-----------------------|---------|---|--------|------------|-----|--------|------|
| Query Match | 100.0%; | Score | 768; | DB | 19; | Length | 151; |
| Best Local Similarity | 100.0%; | Pred. No. | 7e-73; | | | | |
| Matches | 151; | Conservative | 0; | Mismatches | 0; | Indels | 0; |
| Gaps | 0; | | | | | | |
| QY | 1 | MKLLKVAAPAAIVVSGSALAGVFPQWGGGNHNGGNSGPDSTLSIYQGSANAALAQ | 60 | | | | |
| Db | 1 | MKLLKVAAPAAIVVSGSALAGVFPQWGGGNHNGGNSGPDSTLSIYQGSANAALAQ | 60 | | | | |
| QY | 61 | SDARKSETTIQTSGYNGGADVQGGADNSTLTQTQNGFRNNATYDQLVTRVVTWTHMAHAGG | 120 | | | | |
| Db | 61 | SDARKSETTIQTSGYNGGADVQGGADNSTLTQTQNGFRNNATYDQLVTRVVTWTHMAHAGG | 120 | | | | |
| QY | 121 | NNAALVNOTASDSSVMVRQVGFGNNTATANCY | 151 | | | | |
| Db | 121 | NNAALVNOTASDSSVMVRQVGFGNNTATANCY | 151 | | | | |

RESULT 2

```

RES0011 2
US-09-543-407-24
/ Sequence 24, Application US/09543407
/ GENERAL INFORMATION:
/ APPLICANT: White, Aaron P.
/ APPLICANT: Dotan, James L.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Kay, William W.
/ TITLE OF INVENTION: BACTERIAL FILMSRIAL SYSTEM FOR
/ PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
/ FILE REFERENCE: 920043.406
/ CURRENT APPLICATION NUMBER: US/09/543,407
/ CURRENT FILING DATE: 2000-04-05
/ NUMBER OF SEQ ID NOS: 59
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 24
/ LENGTH: 151
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
/ OTHER INFORMATION: sequence containing the replacement fragment
/ OTHER INFORMATION: encoding P73 from GPe3 of Le-shmania major.
US-09-543-407-24

```

| Query Match | 91.1%; | Score 700; | DB 19; | Length 151; |
|-----------------------|-----------------|---------------------|---------------------------------------|----------------|
| Best Local Similarity | 89.9%; | Pred. No. 1.12e-65; | | |
| Matches 143; | Conservative 0; | Mismatches 0; | Indels 16; | Gaps 2; |
| QY | 1 | MKLLKVAAPAAI | VVSGSALAGVVPQGGGGNGNGGSGGPDSTLSII | QYCSNNAALAQ 60 |
| Db | 1 | MKLLKVAAPAAI | VVSGSALAGVVPQGGGGNGNGGSGGPDSTLSII | QYCSNNAALAQ 60 |
| QY | 61 | SDARKSETTITQ | SGYNGADVGGGADNSIIELTQTQGFNNAT----- | YDQLVTRVVT 112 |
| Db | 61 | SDARKSETTITQ | SGYNGADVGGGADNSIIELTQTQGFNNATIOQWAKXV | YDQLVTRVVT 120 |
| QY | 113 | HEMAGAGNNAAL | VNQTASDSSVVMVQVFGGNATANQY | 151 |
| Db | 121 | HEMAHA----- | NOTASDSSVVMVQVFGGNATANQY | 151 |

RESULT 3

```

RESUL 3
US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543, 407

```

```

; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO: 5
; LENGTH: 151
; TYPE: PRN
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

```

| | |
|---------------------------|---|
| Query Match | 88.7%; Score 681; DB 19; Length 151; |
| Best Local Similarity | 90.7%; Pred. No. 1.2e-63; |
| Matches 137; Conservative | 2; Mismatches 12; Indels 0; Gaps 0; |
| QY | 1 MKLIKVAFAAIVVSGSALAGVVPQGGGNNHNGGNSGGPDSLUSIIQYGSANAALQ 60 |
| Db | 1 MKLIKVAFAAIVVSGSALAGVVPQGGGNNHNGGNSGGPDSLUSIIQYGSANAALQ 60 |
| QY | 61 SDARKSETTITQSGYNGADVGGADNSIETLTQGFRRNATYDQIVRVVTHEMAGG 120 |
| Db | 61 SDARKSETTITQSGYNGADVGGADNSIETLTQGFRRNATIDQNAKNSDIIVGQYGG 120 |
| QY | 121 NNAALVNCQTASDSVVMVQVGFGNATANQY 151 |
| Db | 121 NNAALVNCQTASDSVVMVQVGFGNATANQY 151 |

RESULT 4

```

US-08-233-642A-57
/ Sequence 57, Application US/08233642A
/ GENERAL INFORMATION:
/ APPLICANT: Kay, William W.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Clouthier, Sharon C.
/ APPLICANT: Doran, James L.
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
/ TITLE OF INVENTION: BASED VACCINES
/ NUMBER OF SEQUENCES: 58
/ NUMBER OF SEQUENCES: 58
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Seed and Berry
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: U.S.A.
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/233.642A
/ FILING DATE: 26-APR-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: King, Joshua
/ REGISTRATION NUMBER: 35,570
/ REFERENCE/DOCKET NUMBER: 920043.403C3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 692-6031
/ TELEX: 3723836 SEEDANBERRY
/ INFORMATION FOR SEQ ID NO: 57:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 151 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-233-642A-57

```

Query Match 88.0%; Score 676; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 4.2e-63;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYGG 120
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 NNPAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 5

US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-14

Query Match 85.8%; Score 659; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 2.7e-61; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNAT-----YDQ 105
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120
QY 106 LVTRVVTHEMAHAGGNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 121 LVTRVVTHEMAH-----SVVMVRQVGFNNATANQY 151

RESULT 6

US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-22

Query Match 85.5%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.3e-61; Indels 30; Gaps 2;
Matches 136; Conservative 0; Mismatches 0;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAH-- 118
Db 61 SDARKSETTITQSGYNGADVQGGADN-----YDQLVTRVVTHEMAHADQ 105
QY 119 -----GNNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 106 WNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7

US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-28

Query Match 79.8%; Score 613; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2e-56;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAH-- 118
Db 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAHAFR 98
QY 119 -----GNNNAALVNQTASDSSVMVRQVGFNNATANQY 151
Db 99 NNATIDQWNAKNSDITVGQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 8

US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Kay, William W.
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 12
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
 ; OTHER INFORMATION: sequence containing the replacement fragment
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
 US-09-543-407-12

Query Match 78.5%; Score 603; DB 19; Length 151;
 Best Local Similarity 80.8%; Pred. No. 2.4e-55;
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151

RESULT 9
 US-09-543-407-26
 ; Sequence 26, Application US/09543407
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Aaron P.
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Kay, William W.
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 26
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
 ; OTHER INFORMATION: sequence containing the replacement fragment
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
 US-09-543-407-26

Query Match 78.1%; Score 600; DB 19; Length 151;
 Best Local Similarity 81.5%; Pred. No. 4.9e-55;
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALYDQ 60
 QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 LVTRVVTHEMAHAGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151

RESULT 10
 US-09-543-407-18
 ; Sequence 18, Application US/09543407
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Aaron P.
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Kay, William W.
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 18
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
 ; OTHER INFORMATION: sequence containing the replacement fragment
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
 US-09-543-407-18

Query Match 78.0%; Score 599; DB 19; Length 151;
 Best Local Similarity 81.5%; Pred. No. 6.3e-55;
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
 QY 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 DB 1 MKLLKVAFAAIVVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60
 QY 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120
 DB 61 SPARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
 QY 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151
 DB 121 NNAALVNOTASDSSVMVQVGFNNATANQY 151

RESULT 11
 US-09-543-407-20
 ; Sequence 20, Application US/09543407
 ; GENERAL INFORMATION:
 ; APPLICANT: White, Aaron P.
 ; APPLICANT: Doran, James L.
 ; APPLICANT: Collinson, S. Karen
 ; APPLICANT: Kay, William W.
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 ; FILE REFERENCE: 920043.406
 ; CURRENT APPLICATION NUMBER: US/09/543,407
 ; CURRENT FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 20
 ; LENGTH: 151
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
 ; OTHER INFORMATION: sequence containing the replacement fragment
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
 US-09-543-407-20

Query Match 77.7%; Score 597; DB 19; Length 151;

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 73.7%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2e-51;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7; Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 77.3%; Score 594; DB 19; Length 131;
Best Local Similarity 89.3%; Pred. No. 1.8e-54;
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 21 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 VGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGGNNALVNOTASDSSVMVRQV 140
Db 61 VGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGGNNALVNOTASDSSVMVRQV 120
QY 141 GFNNATANQY 151
Db 121 GFNNATANQY 131

RESULT 13
US-09-543-407-16
; Sequence 16; Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 73.7%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 2e-51;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7; Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 67.3%; Score 517; DB 19; Length 151;
Best Local Similarity 69.5%; Pred. No. 3.2e-46;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 61 TDARNSDLTITQGGGNGADVGGGSDSSIDLTQRFNGSATLDDOWNGNKSNMTVQVQFG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NGCAAVDQTASNSVNVTVQVGFNNATANQY 151

RESULT 15
US-08-978-878-4
; Sequence 4; Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match          66.9%; Score 514; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.7e-46;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db       1 MKLLKVAATAIIVFSGSAVAGVVPQYGGGNGHNGGSSGPNSEINIIYQGGNSALALQ 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNNATYDCLVTRVVVTHEMAHAGG 120
        :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db       61 TDARNSDLTITQGGGNGADVCGGSDSSIDLITQGFNGSATLDWNGKNSMTVKQFGG 120
        :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||

QY      121 NNAALVNCTASDSVYVQVGFGNATANQY 151
        |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
Db       121 GNGAAVDQTASNSVNVTVQVGFGNATAHQY 151
        |||  :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
```

Search completed: August 2, 2004, 15:26:47
Job time : 168.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds
(without alignments)
888.146 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:*

1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 514 | 66.9 | 151 | 5 | US-09-741-873C-4 |
| 2 | 436 | 56.8 | 131 | 5 | US-09-741-873C-2 |
| 3 | 101.5 | 13.2 | 520 | 6 | US-10-479-638-21 |
| 4 | 93.5 | 12.2 | 256 | 6 | US-10-425-115-301334 |
| 5 | 93 | 12.1 | 258 | 6 | US-10-425-115-300390 |
| 6 | 92 | 12.0 | 295 | 6 | US-10-425-115-312468 |
| 7 | 91.5 | 11.9 | 179 | 6 | US-10-425-115-346132 |
| 8 | 91.5 | 11.9 | 1203 | 6 | US-10-170-205E-35751 |
| 9 | 91.5 | 11.9 | 1203 | 6 | US-10-170-205E-741 |
| 10 | 91.5 | 11.9 | 1327 | 1 | PCT-US04-02338-49 |
| 11 | 88.5 | 11.5 | 197 | 6 | US-10-425-115-304391 |
| 12 | 88.5 | 11.5 | 388 | 5 | US-09-248-796A-17306 |
| 13 | 87.5 | 11.4 | 573 | 7 | US-60-585-632-7807 |
| 14 | 87.5 | 11.4 | 573 | 7 | US-60-579-062-7907 |
| 15 | 86.5 | 11.3 | 193 | 6 | US-10-425-115-254240 |
| 16 | 86 | 11.2 | 131 | 7 | US-60-565-632-11109 |
| 17 | 86 | 11.2 | 131 | 7 | US-60-579-062-11109 |
| 18 | 85 | 11.1 | 321 | 1 | PCT-US04-05654-590 |
| 19 | 85 | 11.1 | 511 | 6 | US-10-425-115-320950 |
| 20 | 85 | 11.1 | 841 | 7 | US-60-585-632-7906 |
| 21 | 85 | 11.1 | 841 | 7 | US-60-579-062-7906 |
| 22 | 84.5 | 11.0 | 412 | 7 | US-60-565-632-7905 |
| 23 | 84.5 | 11.0 | 412 | 7 | US-60-579-062-7905 |
| 24 | 84 | 10.9 | 556 | 6 | US-10-425-115-337674 |
| 25 | 84 | 10.9 | 586 | 1 | PCT-US03-24982A-317 |
| 26 | 83.5 | 10.9 | 758 | 7 | US-60-565-632-9194 |

ALIGNMENTS

RESULT 1

```

US-09-741-873C-4
; Sequence 4; Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-4

```

```

Query Match 66.9%; Score 514; DB 5; Length 151;
Best Local Similarity 68.9%; Pred. No. 3.7e-37;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLKVAFAAIVVSGSALAAGVVPQGGGNGGNGSSGDPSTLSIYQYGSAALALQ 60
Db 1 MKLKVAFAAIVVSGSALAAGVVPQGGGNGGNGSSGDPSTLSIYQYGSAALALQ 60
QY 61 SPARKSETTITGSGYNGADVGQGDADNSTFIETQGFNRNATYDOLVTRVVTHEMAHAG 120
Db 61 TPARNSDLTITGSGYNGADVGQGDADNSTFIETQGFNRNATYDOLVTRVVTHEMAHAG 120
QY 121 NNAALVNQPADSSVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATANQY 151

```

```

27 83.5 10.9 758 7 US-60-579-062-9194 Sequence 9194, Ap
28 83 10.8 376 6 US-10-491-733-2 Sequence 2, Appli
29 83 10.8 719 5 US-09-248-796A-17559 Sequence 17559, A
30 82.5 10.7 630 7 US-60-581-351-9810 Sequence 9810, Ap
31 82 10.7 279 6 US-10-425-115-343835 Sequence 343835,
32 82 10.7 443 6 US-10-100-683-7608 Sequence 7608, Ap
33 82 10.7 1358 6 US-10-778-804-11 Sequence 11, Appl
34 81.5 10.6 508 6 US-10-425-115-285216 Sequence 285216,
35 81.5 10.6 532 6 US-10-425-115-285214 Sequence 285214,
36 81 10.5 201 6 US-10-425-115-309662 Sequence 309662,
37 81 10.5 376 6 US-10-425-115-342526 Sequence 342526,
38 81 10.5 1127 7 US-60-581-351-1872 Sequence 1872, Ap
39 80.5 10.5 246 6 US-10-854-439-511 Sequence 511, App
40 80.5 10.5 269 6 US-10-425-115-190187 Sequence 190187,
41 80.5 10.5 630 7 US-60-581-351-9806 Sequence 9806, Ap
42 80.5 10.5 631 7 US-60-581-351-9807 Sequence 9807, Ap
43 80.5 10.4 1571 6 US-10-864-138-2 Sequence 2, Appli
44 80 10.4 132 6 US-10-425-115-351875 Sequence 351875,
45 80 10.4 1059 1 PCT-US03-24982A-343 Sequence 343, App

```

```

RESULT 2
US-09-741-873C-2
; Sequence 2, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Olsen, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT FILING DATE: 2000-12-22
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873C-2

```

```

Query Match          56.8%; Score 436; DB 5; Length 131;
Best Local Similarity 65.6%; Pred. No. 1.7e-30;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVVPWGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSETTIQSGYNGAD 80
DB 1 GVVPYGGGNGHNGGNSGPNSELTNYQGGNSALALQTDARNSDTIHQGGNGAD 60
QY 81 VQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGNNAALVNTASDSSVMVRQV 140
DB 61 VQGGDDSSIDLTQRGFNSATLDQWNGKNSMTVKQFGGNGAVIDQTASNSSVNTQV 120
QY 141 GFGNNATANQY 151
DB 121 GFGNNATAHQY 131

```

```

RESULT 3
US-10-479-638-21
; Sequence 21, Application US/10479638
; GENERAL INFORMATION:
; APPLICANT: Don A. Roth
; APPLICANT: Randolph V. Lewis
; APPLICANT: The University of Wyoming
; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
; FILE REFERENCE: WYO-02-0084US
; CURRENT APPLICATION NUMBER: US/10/479,638
; PRIOR FILING DATE: 2003-12-03
; PRIOR APPLICATION NUMBER: PCT/US02/18256
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 60/296,184
; PRIOR FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Argiope trifasciata
US-10-479-638-21

```

```

Query Match          13.2%; Score 101.5; DB 6; Length 520;

```

```

Best Local Similarity 24.6%; Pred. No. 0.68;
Matches 34; Conservative 25; Mismatches 66; Indels 13; Gaps 3;

QY 15 SGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSET-TITQS 73
DB 347 AGAGAAAAASAGAGAGGYGGYGVAGSS-----ISYGATSSSATSSSTRSGIVTSG 402
QY 74 GYNGADVQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGNNAALVNTASDS 133
DB 403 GYGAGAAAGAGACAAAGAGSYGSIIRLSSAEAVNRVSSNICAVASGGASALPG----- 456
QY 134 SVMVRQVGFNNATANQY 151
DB 457 --VISNIFSGVSSAGSY 472

```

```

RESULT 4
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37894C.1.pep
US-10-425-115-301334

```

```

Query Match          12.2%; Score 93.5; DB 6; Length 256;
Best Local Similarity 25.2%; Pred. No. 1.5;
Matches 38; Conservative 15; Mismatches 51; Indels 47; Gaps 6;

QY 17 SALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALQSDA-----RKSETTIQ 72
DB 79 SVVAG-----GGGGGGGGGTGGGGGGGGGTSTAAAGPSSGNYANAEGKAG 133
QY 73 SGYNGAD-----VQGGADNSTIEL--TQNGFRNNATYDQLVTRVVTHEMAHAG 120
DB 134 GGMGGADGAYGAGGAGGYGKGGESGVALAPSSDGYNGGAAD-----ATGGG 182
QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151
DB 183 SGAG-----GGHGGGAGAPSY 198

```

```

RESULT 5
US-10-425-115-300390
; Sequence 300390, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300390
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Zea mays

```

```
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(258)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37025C.1.pep
US-10-425-115-300390

Query Match      12.1%; Score 93; DB 6; Length 258;
Best Local Similarity 26.8%; Pred. No. 1.7; Indels 32; Gaps 5;
Matches 33; Conservative 16; Mismatches 42; Indels 32; Gaps 5;

QY 17 SALAGVPPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG-- 74
DB 78 SSIAG-----GGGGGGGGGCTNGSGSGGGGSGYGSSTTAASGSPSGNYADAEGKGAG 132
QY 75 -----YNGA--DVGGGADNSTIEL--TQGFNNATYDQLVTRVVTTHMAHAGG 120
DB 133 GGMGGGANGAIGSAGGGVGKGEGVGVVALAPSSNGYNGGAAD-----ATGGG 181
QY 121 NNA 123
DB 182 SGA 184

RESULT 6
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-425-115-312468

Query Match      12.0%; Score 92; DB 6; Length 295;
Best Local Similarity 30.2%; Pred. No. 2.4; Indels 22; Gaps 5;
Matches 35; Conservative 13; Mismatches 46; Indels 22; Gaps 5;

QY 26 WGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYGNADVCGGA 85
DB 138 YGGGYSGGGYSSG-GYAANGYGVGSGGNYSNASGGYSGS-----DGYGNGAASGGYA 192
QY 86 DNSTIELTQGFNNATYDQLVTRVVTTHMAHAGGNN-----AALVNQTASDSS 134
DB 193 NN-----LSSGYNSGRYN-----TIGSDGNTGGYNSYENPYGAGNYNTGGSSSS 238

RESULT 7
US-10-425-115-346132
; Sequence 346132, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346132
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pep
US-10-425-115-346132

Query Match      11.9%; Score 91.5; DB 6; Length 179;
Best Local Similarity 27.7%; Pred. No. 1.5; Indels 37; Gaps 6;
Matches 39; Conservative 10; Mismatches 55; Indels 37; Gaps 6;

QY 19 LAGVPPQWGGG-----NHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTI 70
DB 64 LDGLLSLGGGLDGLLGLTGGGGKKNQAADSGNAQEGSGNA---QEDSGNAQ--- 116
QY 71 TQSGYGNADVGGGADNSTIELTQGFNNATYDQLVTRVVTTHMAHAGGNNALVN-QT 129
DB 117 EESGNNQNAAGAGAGA-----ENGANNGT-----EAAGAENAAAGNGQA 155
QY 130 ASDSSVMVRQVGVGNNTANQ 150
DB 156 QIEGEDEQAQGNAGNENAAEE 176

RESULT 8
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Clone ID: MRT4577_48027C.1.pep
US-10-170-205E-35751

Query Match      11.9%; Score 91.5; DB 6; Length 299;
Best Local Similarity 29.4%; Pred. No. 2.6; Indels 19; Gaps 6;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNSNPSSSSPTSS-SSSSPSPSGSLAESPEAA 157
QY 65 KSETTIT---QSGYGNADVCGGADNSTIELTQGFNNATYDQLVTRVVTTHMAHAGG 120
DB 158 GVSSTAPLQFGAAGPGTGVPVAVSGALRELLAACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
DB 207 VNAAK--DMAGRKSSPLHFAAGFG 227

RESULT 9
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patent in version 3.2
```


Db 240 GAGFGDNTSSY 252

RESULT 13

US-60-565-632-7907

Sequence 7907, Application US/60565632

GENERAL INFORMATION:

APPLICANT: Monsanto Technology, LLC

APPLICANT: Baum, James A

APPLICANT: Kovalic, David K.

APPLICANT: Larosa, Thomas J.

APPLICANT: Lu, Maolong

APPLICANT: Munyikwa, Tichifa R. I.

APPLICANT: Roberts, James K.

APPLICANT: Wu, Wei

APPLICANT: Zhang, Bei

TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

TITLE OF INVENTION: Compositions Thereof

FILE REFERENCE: 38-21(53403)B

CURRENT APPLICATION NUMBER: US/60/565,632

CURRENT FILING DATE: 2004-04-27

NUMBER OF SEQ ID NOS: 15449

SOFTWARE: PatentIn version 3.2

SEQ ID NO 7907

LENGTH: 573

TYPE: PRT

ORGANISM: Diabrotica virgifera

US-60-565-632-7907

Query Match 11.4%; Score 87.5; DB 7; Length 573;

Best Local Similarity 26.0%; Pred. No. 12;

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;

QY 30 GNHNGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSGYNGAD 80

Db 241 GNENGTGAENNADADTDVAQ--GSTNEAENNANADVQNDAAQANENGAAENSGNAD 299

QY 81 VGQGAN--STIELTQN-----GFRNNATYDQLVTRVVTHEMAHAGGNNNA 123

Db 300 AAQGTNDGAAAGTGNADPAQGNNDNGAAENSGNENGTAENNANADVQNDAAQVNDNGA 359

QY 124 ALVNQTSADSS 134

Db 360 AENNGNADAA 370

RESULT 14

US-60-579-062-7907

Sequence 7907, Application US/60579062

GENERAL INFORMATION:

APPLICANT: Baum, James A

APPLICANT: Kovalic, David K

APPLICANT: Larosa, Thomas J

APPLICANT: Lu, Maolong

APPLICANT: Munyikwa, Tichifa R. I.

APPLICANT: Roberts, James K

APPLICANT: Wu, Wei

APPLICANT: Zhang, Bei

TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

TITLE OF INVENTION: Compositions thereof

FILE REFERENCE: 38-21 (53403) C

CURRENT APPLICATION NUMBER: US/60/579,062

CURRENT FILING DATE: 2004-06-11

NUMBER OF SEQ ID NOS: 41445

SOFTWARE: PatentIn version 3.2

SEQ ID NO 7907

LENGTH: 573

TYPE: PRT

ORGANISM: Diabrotica virgifera

US-60-579-062-7907

Query Match 11.4%; Score 87.5; DB 7; Length 573;

Best Local Similarity 26.0%; Pred. No. 12;

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;

QY 30 GNHNGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSGYNGAD 80

Db 241 GNENGTGAENNADADTDVAQ--GSTNEAENNANADVQNDAAQANENGAAENSGNAD 299

QY 81 VGQGAN--STIELTQN-----GFRNNATYDQLVTRVVTHEMAHAGGNNNA 123

Db 300 AAQGTNDGAAAGTGNADPAQGNNDNGAAENSGNENGTAENNANADVQNDAAQVNDNGA 359

QY 124 ALVNQTSADSS 134

Db 360 AENNGNADAA 370

RESULT 15

US-10-425-115-254240

Sequence 254240, Application US/10425115

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants

FILE REFERENCE: 38-21(53222)B

CURRENT APPLICATION NUMBER: US/10/425,115

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 369326

SEQ ID NO 254240

LENGTH: 193

TYPE: PRT

ORGANISM: Zea mays

FEATURE:

OTHER INFORMATION: Clone ID: MRT4577_163446C.1.pap

US-10-425-115-254240

Query Match 11.3%; Score 86.5; DB 6; Length 193;

Best Local Similarity 26.9%; Pred. No. 4.3;

Matches 35; Conservative 15; Mismatches 53; Indels 27; Gaps 5;

QY 27 GCGGNHNGG---GNSGSPDSTLSIYQGSANAALAALQSDARKSETTITQSGYNGADYQ 83

Db 76 GGGASQNGSGYSGSGSGSGSSTYSQGGYSGYGSNA-----GGTGGGGGGGQ 126

QY 84 --GADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNNNAALVNQT-----ASDSS 134

Db 127 AGGYWNSNAQSGSGTSGSSYANRYWDCSSEGGANANGGGMNSQNGGGGGGGGSDGS 186

QY 135 VMVQVGFQGN 144

Db 187 -----GYGN 190

Search completed: August 2, 2004, 15:29:56

Job time : 18.8 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds
(without alignments)
1545.204 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 681 | 88.7 | 151 | 2 JC6039 | fimbrin protein ag |
| 2 | 681 | 88.7 | 151 | 2 A10635 | major curlin chain |
| 3 | 517 | 67.3 | 151 | 2 S70788 | curlin protein csg |
| 4 | 502.5 | 65.4 | 152 | 2 D90806 | curlin major subun |
| 5 | 502.5 | 65.4 | 152 | 2 H85665 | hypothetical prote |
| 6 | 104.5 | 13.6 | 151 | 2 S70787 | curlin nucleator p |
| 7 | 104.5 | 13.6 | 151 | 2 C90806 | minor curlin subun |
| 8 | 104.5 | 13.6 | 151 | 2 G85665 | curlin minor chain |
| 9 | 101 | 13.2 | 1748 | 2 S42136 | cnjB protein - Tet |
| 10 | 99.5 | 13.0 | 151 | 2 JC6040 | fimbrin protein ag |
| 11 | 99.5 | 13.0 | 151 | 2 AH0635 | nucleation compone |
| 12 | 99.5 | 13.0 | 2174 | 2 E95965 | hypothetical glyci |
| 13 | 98 | 12.8 | 145 | 2 AD3143 | conserved hypohet |
| 14 | 98 | 12.8 | 145 | 2 H98144 | hypothetical prote |
| 15 | 97.5 | 12.7 | 645 | 2 F70825 | probable PPE prote |
| 16 | 96.5 | 12.6 | 313 | 2 T04776 | hypothetical prote |
| 17 | 94 | 12.2 | 354 | 2 B70663 | probable PPE prote |
| 18 | 94 | 12.2 | 656 | 2 D96831 | hypothetical prote |
| 19 | 93.5 | 12.2 | 256 | 2 T03371 | glycine-rich prote |
| 20 | 93.5 | 12.2 | 2204 | 2 A70534 | probable PPE prote |
| 21 | 92.5 | 12.0 | 1028 | 2 A56038 | DNA-binding protei |
| 22 | 92.5 | 12.0 | 1213 | 2 S16356 | ovo protein - fru1 |
| 23 | 92 | 12.0 | 575 | 2 S35327 | protein kinase sgg |
| 24 | 91.5 | 11.9 | 409 | 2 T20847 | hypothetical prote |
| 25 | 91.5 | 11.9 | 573 | 2 C86266 | F3f19.21 protein - |
| 26 | 91.5 | 11.9 | 1910 | 2 AF0394 | probable adhesin h |
| 27 | 91 | 11.8 | 652 | 2 E97857 | cell surface antig |
| 28 | 90.5 | 11.8 | 447 | 2 G84657 | probable disease r |
| 29 | 89.5 | 11.7 | 963 | 2 B70524 | probable PPE prote |

| | | | | | |
|----|------|------|------|----------|--------------------|
| 30 | 89.5 | 11.7 | 1238 | 2 AH0038 | probable exported |
| 31 | 89 | 11.6 | 343 | 2 T05221 | hypothetical prote |
| 32 | 89 | 11.6 | 959 | 2 B44402 | nuclear pore compl |
| 33 | 88.5 | 11.5 | 967 | 2 S66852 | hypothetical prote |
| 34 | 88.5 | 11.5 | 1217 | 2 S52714 | sericin1B - silkwo |
| 35 | 88 | 11.5 | 407 | 2 T21956 | hypothetical prote |
| 36 | 88 | 11.5 | 586 | 2 T26667 | hypothetical prote |
| 37 | 88 | 11.5 | 1778 | 2 T50074 | probable nucleopor |
| 38 | 87.5 | 11.4 | 1053 | 2 B70987 | probable PPE prote |
| 39 | 87 | 11.3 | 3624 | 2 E70946 | probable PPE prote |
| 40 | 87 | 11.3 | 262 | 2 AD0835 | large repetitive p |
| 41 | 86.5 | 11.3 | 262 | 2 S00275 | tail fiber protein |
| 42 | 86.5 | 11.3 | 590 | 1 A45621 | leishmanolysin (EC |
| 43 | 86.5 | 11.3 | 745 | 2 E64559 | outer membrane pro |
| 44 | 86 | 11.2 | 582 | 2 F70675 | probable PPE prote |
| 45 | 85 | 11.2 | 639 | 2 C42049 | leishmanolysin (EC |

ALIGNMENTS

RESULT 1

JC6039
fimbrin protein agfa precursor - Salmonella enteritidis
C/Species: Salmonella enteritidis
C/Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text_change 08-Oct-1999
C/Accession: JC6039; PC6015; A44898
R/Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A/Title: Salmonella enteritidis agfA operon encoding thin, aggregative fimbriae.
A/Reference number: JC6039; MUID:96146512; PMID:8550497
A/Accession: JC6039
A/Molecule type: DNA
A/Residues: 1-151 <COL>
A/Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43599.1; PID:gl184714
A/Accession: PC6015
A/Molecule type: Protein
A/Residues: 21-52 <CO2>
A/Experimental source: strain 27655-3b
A/Note: the authors translated the codon ACG for residue 44 as Ile
R/Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.
J. Bacteriol. 173, 4773-4781, 1991
A/Title: Purification and characterization of thin, aggregative fimbriae from Salmonella
A/Reference number: A44898; MUID:91310586; PMID:1677357
A/Contents: 27655
A/Accession: A44898
A/Status: preliminary
A/Molecule type: protein
A/Residues: 21-33 <CO3>
A/Note: sequence extracted from NCBI backbone (NCBIP:45936)
C/Genetics:
A/Gene: agfA
C/Function:
A/Description: major component of thin aggregative fimbriae
A/Key: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C/Keywords: fimbria
P/1-20/Domain: signal sequence #status predicted <SIG>
F/1-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 88.7%; Score 681; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.3e-49;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQGSANAALAQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQGSANAALAQ 60

QY 61 SPARKSETTITQSGYNGADVQGGADNSTIETQTQGFNNATYDOLVTRVVTHEMAHAGG 120
Db 61 SPARKSETTITQSGYNGADVQGGADNSTIETQTQGFNNATIDQWNAKNSDIIVQYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2

A:0635
major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: A10635
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: A10635
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176
C:Genetics:
A:Gene: STY1181

Query Match 88.7%; Score 681; DB 2; Length 151;
Best Local Similarity 90.7%; Pred. No. 3.3e-49;
Matches 137; Conservative 2; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVVTHEMAHAGG 120
Db 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVVTHEMAHAGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788
curlin protein csgA precursor - Escherichia coli (strain K-12)
N:Alternate names: csgA protein; major curlin protein
C:Species: Escherichia coli
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70788; G64846; S34560; S34559
R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and CbpA
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70788
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64846
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;
A:Experimental source: strain K-12, substrain MG1655
R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.
Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA
A:Reference number: S31202; MUID:93211294; PMID:8459772
A:Accession: S31202
A:Molecule type: DNA
A:Residues: 1-6 'V', 8-151 <OLS1>
A:Cross-references: EMBL:L04979
A:Accession: S34560
A:Molecule type: protein
A:Residues: 21-42; 44-50 <OLS2>
R:Olsen, A.N.; Arngqvist, A.M.
submitted to the EMBL Data Library, October 1992
A:Reference number: S34559
A:Accession: S34559
A:Molecule type: DNA
A:Residues: 1-133; 'RQDSGMLW' <OLS3>
A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425
A:Experimental source: strain K-12, substrain W3110
C:Genetics:
A:Gene: csgA
A:Map position: 23.15
C:Function:
A:Description: major component of wild-type curli; interaction between CsgA and CsgB tra
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F;A-20/Domain: signal sequence #status predicted <SIG>
F;21-151/Product: curlin #status experimental <MAT>

Query Match 67.3%; Score 517; DB 2; Length 151;
Best Local Similarity 69.5%; Pred. No. 1e-35;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVVTHEMAHAGG 120
Db 61 TDARNSDLTITQHGNGGADVCGGSDSDSIDITQFGNSATLDQNGKNSMTVQYQGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGAAYDQASNSVNVTVQVGFNNATAHQY 151

RESULT 4

D90806
curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD
D90806)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: D90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90806
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: Ecs1420

Query Match 65.4%; Score 502.5; DB 2; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.6e-34;
Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAAPAAIIVSGSALAGVVPQYGGGNGHNGSGSPDSTLSIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATYDQLVTRVVTHEMAHAG 119

[illegible]

Db 74 EGENNRAKVDQAGNVNFAY-IEQTGNANDASIGSAYGNSAAIIQKSGNKANITQY 129

RESULT 12

E95965
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) magad

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: E95965

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: E95965

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-2174 <XLR>

A:Cross-references: GB:AL591985; PIDN:CA049389.1; PID:G15140875; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; A

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kaïman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaült, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMB21548

A:Genome: plasmid

| | | | | |
|-----------------------|------------------|--|-----------------------------|--------------|
| Query Match | 13.0%; | Score 99.5; | DB 2; | Length 2174; |
| Best Local Similarity | 23.0%; | Pred. No. 4.3; | | |
| Matches 32; | Conservative 14; | Mismatches 66; | Indels 27; | Gaps 2; |
| QY | 11 | AIIVSGSALAGVFPQWGGGHHGGKSSGGPDS | TLISIIYQGSANAALAQSDARKSETTI | 70 |
| Db | 892 | AVFTKGS | SHLLAQS | IGGGGNGGSASS |
| QY | 71 | TQSGYGNAGADVQCGADNSTTIELTQNGFRNNATYDQLVTRVVTHMAHAGGNAALVNQTA | 130 | |
| Db | 933 | ADIGSVVSGSGSAGNGGAVTAALKDSASVT | TADDYANAIVQSIGGGGGNGGV | 986 |
| QY | 131 | SDGSVMVRQVGFGNNTAN | 149 | |
| Db | 987 | --GSVNSKEIGSGFNLTAN | 1003 | |

RESULT 13
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3143
R:Wood, D.W.; Serubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClelland, Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE008699; PIDN:AL4562.1; FID:gi17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

```

Query Match      12.8%; Score 98; DB 2; Length 145;
Best Local Similarity 27.1%; Pred. No. 0.29;
Matches 38; Conservative 25; Mismatches 57; Indels 20; Gaps 5

QY 3 LKVAFAAIVVSGSALAGVVP-----QWG-----GGNHNGGN-----SSGPDST 44
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYQWHSNSAGGAQEGYGNRIPTYQNGGYNR 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 LSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATVD 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 IVGHQYGRHNLS-AVGQEGHDNYGSETTQGNRRNVAGIQFGSNGHTTILTDGNGNIAAGV 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 105 QLVTRVVTHEMAHAGGNAA 124
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 Q-VGRGCSANVSQGGDNVA 138
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
H98144
hypothetical protein AGR_L228 [imported] - Agrobacterium tumefaciens (strain
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: H98144
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.;
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Ma
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobac
A/Reference number: A97359; MUID:21608551; PMID:11743194

```

A:Accession: H58144
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-145 <Kur>
A:Cross-references: GB:AE007870; PID:G15158413; GSPDB:GN00170
C:Genetics:
A:Gene: AGR L 228
A:Map position: linear chromosome

| | | | | |
|-----------------------|-------|--|------------------------|----------------|
| Query Match | 12.8% | Score 98 | DB 2 | Length 145 |
| Best Local Similarity | 27.1% | Pred.No. 0.29 | | |
| Matches | 38 | Conservative 25 | Mismatches 57 | Indels 20 |
| Gaps | 5 | | | |
| Qy | 3 | LLKVAAPAAIVSSGALAGVVP | -----QWG-----GGGNHNGGN | -----SSGPST 44 |
| Db | 1 | MIKSFSTASLVALVGLSAAAPANVDRIEQWMSAGAQEGYGNRIITYQNGGYNR | | |
| Qy | 45 | LSIYQYGSANAALAQDARKSETITQSGYNGADVQGGADNFTIELTQNGFRNATYD | | |
| Db | 61 | IVGHQYGRHLS AVGQEGHDNYGSTTQNGRNRVAGTQFQSNHTTILTDGNGNIAQV | | |
| Qy | 105 | QLTVRVVTHMAHAGGNAA | | 124 |
| Db | 120 | Q-VGRGCSANVSGNDNVA | | 138 |

RESULT 15
F70825.
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70825
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Connor, R. A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Ho-
Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998
A: Authors, Sgares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.
A: Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: F70825
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-645 <COL>
A: Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PID:CAA17522.1; P1

A/Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17522.1; PID:e125329

Tue Aug 3 10:54:46 2004

A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 12.7%; Score 97.5; DB 2; Length 645;
Best Local Similarity 25.9%; Pred. No. 1.6;
Matches 43; Conservative 15; Mismatches 59; Indels 49; Gaps 9;

QY 15 SGSALAGVWPQ-----WGGGNNHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETT 69
DB 237 SGNVVGIGPSSFNVSGNIGNANVGGNSG-DNNFGNFGNANIGIGNAGPNMSSPAV 295
QY 70 ITQS-----GYNGADVGGADNSTIELTON-----GFRNATYDQLVTRVVT 112
DB 296 PTPGNVVGIGNGGNGNFGGN-----TGNANIGLVNVDGNVGFNGSGSYN----- 342
QY 113 HEMAHAGGNAAL----VNQTA-----SDSSVMVRQVGFGNNTAN 149
DB 343 FGFGTGNNNIGLGTGSNQIGFGLNSGSG----NIGFGNSGTGN 384

Search completed: August 2, 2004, 14:56:26
Job time : 10.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds
(without alignments)
1483.508 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRFQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 681 | 88.7 | 151 | 1 | CSGA_SALTY |
| 2 | 517 | 67.3 | 151 | 1 | CSGA_ECOLI |
| 3 | 502.5 | 65.4 | 152 | 1 | CSGA_ECO57 |
| 4 | 104.5 | 13.6 | 151 | 1 | CSGB_ECOLI |
| 5 | 99.5 | 13.0 | 151 | 1 | CSGB_SALTY |
| 6 | 99.5 | 13.0 | 151 | 1 | CSGB_SALTY |
| 7 | 92.5 | 12.0 | 1028 | 1 | OVO_DROME |
| 8 | 91.5 | 11.9 | 1327 | 1 | TNKK1_HUMAN |
| 9 | 90 | 11.7 | 1656 | 1 | OMPBB_RICJA |
| 10 | 89 | 11.6 | 959 | 1 | NI00_YEAST |
| 11 | 88 | 11.5 | 1778 | 1 | N189_SCHPO |
| 12 | 86.5 | 11.3 | 262 | 1 | V338_BPT2 |
| 13 | 86.5 | 11.3 | 590 | 1 | GP63_LEIDO |
| 14 | 86 | 11.2 | 646 | 1 | GP63_LEIME |
| 15 | 85.5 | 11.1 | 1567 | 1 | ICEN_XANCT |
| 16 | 85 | 11.1 | 599 | 1 | GP63_LEICH |
| 17 | 85 | 11.1 | 602 | 1 | GP63_LEIMA |
| 18 | 85 | 11.1 | 850 | 1 | PRSN_PIERA |
| 19 | 85 | 11.1 | 1655 | 1 | OMPBB_RICCN |
| 20 | 84.5 | 11.0 | 678 | 1 | YF48_MYCTU |
| 21 | 84.5 | 11.0 | 760 | 1 | YBIL_ECOLI |
| 22 | 84.5 | 11.0 | 1196 | 1 | ICBV_PSRSX |
| 23 | 84.5 | 11.0 | 1469 | 1 | ICLV_DROME |
| 24 | 84 | 10.9 | 363 | 1 | ONPC_KLEBN |
| 25 | 84 | 10.9 | 576 | 1 | ICAF_DROME |
| 26 | 83.5 | 10.9 | 1148 | 1 | ICER_PSRSX |
| 27 | 83 | 10.8 | 401 | 1 | YK03_CABEL |
| 28 | 82.5 | 10.7 | 592 | 1 | CEA_CITFR |
| 29 | 81.5 | 10.6 | 172 | 1 | CH18_DROME |
| 30 | 81.5 | 10.6 | 347 | 1 | MSA2_PLAF2 |
| 31 | 81.5 | 10.6 | 392 | 1 | HWE1_HUMAN |
| 32 | 81 | 10.5 | 165 | 1 | GRPI_ORYSA |
| 33 | 81 | 10.5 | 1845 | 1 | Z236_HUMAN |

| | | | | | |
|----|------|------|------|---|------------|
| 34 | 80.5 | 10.5 | 534 | 1 | GCR2_YEAST |
| 35 | 80.5 | 10.5 | 874 | 1 | AL36_AZOVI |
| 36 | 80.5 | 10.5 | 880 | 1 | GUN4_THEFU |
| 37 | 80.5 | 10.5 | 1210 | 1 | ICEN_PSEFL |
| 38 | 80.5 | 10.5 | 1258 | 1 | ICEN_ERWHE |
| 39 | 80 | 10.4 | 380 | 1 | PLXA_COLGL |
| 40 | 80 | 10.4 | 491 | 1 | YK98_MYCTU |
| 41 | 80 | 10.4 | 850 | 1 | PRSN_PIEBR |
| 42 | 80 | 10.4 | 1571 | 1 | C3G_DROME |
| 43 | 80 | 10.4 | 5560 | 1 | SPEN_DROME |
| 44 | 79.5 | 10.4 | 163 | 1 | HCI_NATPH |
| 45 | 79.5 | 10.4 | 663 | 1 | DUS8_MOUSE |

ALIGNMENTS

RESULT 1
CSGA_SALTY STANDARD; PRT; 151 AA.
AC P55225;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Major curlin subunit precursor (Fimbrin SEF17).
GN CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.
OS Salmonella typhimurium,
OS Salmonella typhi, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602, 601, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
[2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
[3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dow L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
[4]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.P., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2
 RL and CT18.";
 RN J. Bacteriol. 185:2330-2337(2003).
 RP [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
 RL fimbriae";
 RN J. Bacteriol. 178:662-667(1996).
 RP [6]
 RP SEQUENCE OF 21-151 FROM N.A.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=94013373; PubMed=8104955;
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,
 RP Munro C.K., Kay C.M., Bansen P.A., Peterkin P.I., Kay W.W.;
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,
 RL the structural gene for thin, aggregative fimbriae.";
 RN J. Clin. Microbiol. 31:2263-2273(1993).
 RP [7]
 RP SEQUENCE OF 21-33.
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;
 RX MEDLINE=91310586; PubMed=1677357;
 RA Collinson S.K., Emsedy L., Mueller K.-M., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RL *Salmonella enteritidis*.";
 RN J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ002301; CAA05317.1; -;
 DR EMBL; AE008749; AAL20074.1; -;
 DR EMBL; AL627269; CAD08268.1; -;
 DR EMBL; AE016840; AAO89399.1; -;
 DR EMBL; U43280; AAC43599.1; -;
 DR PIR; JC6039; JC6039.
 DR StyGene; SG10608; CSGA.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 151
 FT CONFLICT 134 151
 FT REF. 6).
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;
 QY
 Db
 QY
 Db
 QY
 Db
 QY
 Db

Query Match
 Best Local Similarity 88.7%; Score 681; DB 1; Length 151;
 Matches 137; Conservative 12; Mismatches 12; Indels 0; Gaps 0;
 1 MKLLKVAFAAIVSGSALAGVVPWGGGHHNGGNSGPDSTLSIYQYGSAAALQ 60
 1 MKLLKVAFAAIVSGSALAGVVPWGGGHHNGGNSGPDSTLSIYQYGSAAALQ 60
 61 SPARKSETTITQSGYNGGADVQGGADNSTIETQNGFRNATYDQVTRVVVTHEMAHAG 120
 61 SPARKSETTITQSGYNGGADVQGGADNSTIETQNGFRNATYDQVTRVVVTHEMAHAG 120
 121 NNAALVNQATSDSSVWVRQVGFNNATANYQ 151
 121 NNAALVNQATSDSSVWVRQVGFNNATANYQ 151

RESULT 2
 CSGA_ECOLI
 ID _CSGA_ECOLI STANDARD; PRT; 151 AA.
 AC P28307;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major curlin subunit precursor.
 GN CSGA OR B1042.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=932111294; PubMed=8459772;
 RA Olsen A., Arngvist A.;
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional
 RL repression of *csga*, the subunit gene of fibronectin-binding curli in
 RL *Escherichia coli*.";
 RN Mol. Microbiol. 7:523-536(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC4100;
 RX MEDLINE=96414468; PubMed=8817489;
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
 RT "Expression of two *csg* operons is required for production of
 RL fibronectin- and Congo red-binding curli polymers in *Escherichia coli*
 RL K-12.";
 RN Mol. Microbiol. 18:661-670(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1232-1244(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 RL corresponding to the 12.7-28.0 min region on the linkage map.";
 RN DNA Res. 3:137-155(1996).
 RN [5]
 RP SEQUENCE OF 21-40.
 RC STRAIN=K12 / YNML;
 RX MEDLINE=93023873; PubMed=1357528;
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;
 RT "The *Crl* protein activates cryptic genes for curli formation and
 RL fibronectin binding in *Escherichia coli* HB101.";
 RN Mol. Microbiol. 6:2443-2452(1992).
 RN [6]
 RP SEQUENCE OF 21-31.
 RC MEDLINE=91310586; PubMed=1677357;
 RX Collinson S.K., Emsedy L., Trust T.J., Kay W.W.;
 RT "Purification and characterization of thin, aggregative fimbriae from
 RL *Salmonella enteritidis*.";
 RN J. Bacteriol. 173:4773-4781(1991).
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN.

```
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04979; AAA23616.1; -
DR EMBL; X90754; CAA62282.1; -
DR EMBL; AE000205; AAC74126.1; -
DR EMBL; D90741; BAA35832.1; -
DR EMBL; D90742; BAA35840.1; -
DR PIR; S70788; S70788.
DR Ecogen; Egl1489; csGA.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
FT CHAIN 21 151 A -> E (IN REF. 1).
FT CONFLICT 7 7
SQ SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
Query Match 67.3%; Score 517; DB 1; Length 151;
Best Local Similarity 69.5%; Pred. No. 9.7e-37;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGHNGSGSPDSTLSIYQYGSNAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGNPNSELIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSITELTQNGFRNNATYDQLVTRVVTTHMAHAG 120
DB 61 TDARNSDLTITQHGCGNGADVGQSDSDSIDLTQRFNGSATLDQWNGKNSMTVKQFG 120
QY 121 NNAALVNOTASDSSVMVROVGFNNATANY 151
DB 121 GNGAAVDQTASNSVNTVQVGFNNATAHQY 151
RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RT "Mutations in the csGP promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RL Appl. Environ. Microbiol. 67:2367-2370(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP275733; AAK53212.1; -
DR EMBL; AE005315; AAG55788.1; -
DR EMBL; AP002554; BAB34843.1; -
DR PIR; D90806; D90806.
DR PIR; H85665; H85665.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
FT CHAIN 21 152
SQ SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
Query Match 65.4%; Score 502.5; DB 1; Length 152;
Best Local Similarity 68.4%; Pred. No. 1.6e-35;
Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGNGHNGSGSPDSTLSIYQYGSNAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHNGSGNPNSELIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVGQADNSITELTQNGFRNNATYDQLVTRVVTTHMAHAG 119
DB 61 QADARNSDLTITQHGCGNGADVGQSDSDSIDLTQRFNGSATLDQWNGKNSMTVKQFG 120
QY 120 GNAALVNOTASDSSVMVROVGFNNATANY 151
DB 121 GNGAAVDQTASNSVNTVQVGFNNATAHQY 152
RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
ID CSGB_ECOLI
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csG operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli
```

RT K-12."; Mol. Microbiol. 18:661-670(1995).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=K12 / MG1655;

RC MEDLINE=97426617; PubMed=9278503;

RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12.";

RT Science 277:1453-1474(1997).

RL [3]

RN SEQUENCE FROM N.A.

RP STRAIN=K12;

RC MEDLINE=97061202; PubMed=8905232;

RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N., Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M., Horiuchi T.;

RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

RT DNA Res. 3:137-155(1996).

RL [4]

RN SEQUENCE FROM N.A.

RP STRAIN=O157:H7 / EDL933 / ATCC 700927;

RC MEDLINE=21074935; PubMed=11206551;

RX Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RT Nature 409:529-533(2001).

RL [5]

RN SEQUENCE FROM N.A.

RP STRAIN=O157:H7 / RIMD 0509952;

RC MEDLINE=21156231; PubMed=11258796;

RX Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RA "Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";

RT DNA Res. 8:11-22(2001).

RL [6]

RN SEQUENCE OF 1-21 FROM N.A.

RP STRAIN=K12;

RC MEDLINE=95157246; PubMed=7854117;

RX Arngvist A., Olsen A., Normark S.;

RA "Sigma S-dependent growth-phase induction of the csfBA promoter in Escherichia coli can be achieved in vivo by sigma 70 in the absence of the nucleoid-associated protein H-NS.";

RT Mol. Microbiol. 13:1021-1032(1994).

RL [7]

RN COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X90754; CAA62281.1; -.

DR

DR EMBL; AE000205; AAC74125.1; -.

DR EMBL; D90741; BAA35831.1; -.

DR EMBL; AE005315; AAG55787.1; -.

DR EMBL; AP002554; BAB34842.1; -.

DR PIR; C90806; C90806.

DR PIR; G85665; G85665.

DR PIR; S70787; S70787.

DR EcoGene; E012821; cs98.

DR Fimbrin; Signal; Complete proteome.

KW SIGNAL 1 21

FT CHAIN 22 151

FT CHAIN 22 151

SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match 13.6%; Score 104.5; DB 1; Length 151;

Best Local Similarity 29.1%; Pred No. 0.033; 57; Indels 11; Gaps 3;

Matches 34; Conservative 15; Mismatches 15;

QY 38 SSGPSTLSIYQYGSANAALALQSDARKSETTITQSGYGNADVCGQADNSTIELTQNGF 97

Db 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGQATGNSAQLRQGGSKLLAVVAQEGS 76

QY 98 RNNATYDQLVTRVVTHEVAH---AGNNAALVNQASDSSVMVROVGFNNATANQY 151

Db 77 SNRAKIDQ----TGDYNLAIDQAGSANDASISQAGYGNATMIQKSGNKANITQY 129

RESULT 5

CSGB_SALTI STANDARD; PRT; 151 AA.

ID CSGB_SALTI

AC Q827M3; 2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Minor curlin subunit precursor.

GN CSGB OR STV1180 OR T1777.

OS Salmonella typhi.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_TaxID=601;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J., Churcher C., Mangall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;

RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium CT18.";

RT Nature 413:848-852(2001).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=Ty2 / ATCC 700931;

RX MEDLINE=22531367; PubMed=12644504;

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodyouranni V., Schwartz D.C., Blattner F.R.;

RA "Comparative genomics of Salmonella enterica serovar typhi strains Ty2 and CT18.";

RT J. Bacteriol. 185:2330-2337(2003).

RL [3]

RN COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X90754; CAA62281.1; -.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

| | |
|----|-------------------------------------|
| CC | ----- |
| DR | EWEL; AL627269; CAD08267.1; -. |
| DR | EWEL; AE016840; AA069400.1; -. |
| KW | Fimbria; Signal; Complete proteome. |

| | | |
|-------------|---------|-----------------------------------|
| | 1 | POTENTIAL. |
| FT CHAIN | 21 | MINOR CURLIN SUBUNIT. |
| FT SIGNAL | 22 | |
| SQ SEQUENCE | 151 AA; | 16254 MW; 161C54326E573495 CRC64; |

Query Match 13.0%; Score 99.5; DB 1; Length 151;
Best Local Similarity 28.2%; Pred. No. 0.088;
Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

```

Qy      51  GSANAALALQSDARKSE-----TTITQSGYNGADVGG--GADNST-----IELTQ  94
      14  :                               :
      14  GAGGIATATNYDLARSEYFANVELSKSSFNQAAIIIGGVGTGNSARVRCQESKLLSVISQ  73
      :                               :

```

Qy 95 NGFNNATYDQLVTRVVTHTEMAHAGGNNAAALVNQTASDSSWMVRQVGFGNNTANQY 151
| | | | : | | | : | | | : | | |
Dd 74 EGENNRKVDQAGNYNFAY-IETGTANDASISOSAYGNSAALIORGGNKANITQY 129
| | | | | | | | | | | | | | |

RESULT 6
CSGB SALTY

| ID | CSGB_SALTY | STANDARD; | PXT; | 151 AA. |
|----|-------------|---------------------------------|------|---------|
| AC | P55236; | | | |
| DT | 01-OCT-1996 | (Rel. 34, Created) | | |
| DT | 01-OCT-1996 | (Rel. 34, Last sequence update) | | |

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS *Salmonella typhimurium* and

OS *Salmonella typhimurium*, and
OS *Salmonella enteritidis*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OV NCBI TaxID-500 502.
OV NCBI TaxID-500 502.

[1]
 SEQUENCE FROM N.A.
 SPECIES=S.typhimurium; STRAIN=SR-11;
 NCBI_taxid=802, 392;

RA MEDLINE#98111038; PUBMED#9457880;
RA Romling U., Bian Z., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between *Salmonella typhimurium* and
RT *Escherichia coli* with respect to operon structure and regulation."; RT
RT *Escherichia coli* 5598-200. PMID: 9457880.

U. Bacteriol. 180:722-731(1998).
[2]
RN SEQUENCE FROM N.A.
RP
RC SPECIES-S.typhimurium; STRAIN-LT2 / SSSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterstone R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 RT LT2.";

RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N. A.
RC SPECIES=enteritidis; STRAIN=27655-3B;

RX MEDLINE=96146512; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae.";

RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC

CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -|- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

| | |
|----|--------------------------------|
| CC | EMBL; AJ002301; CAA05316.1; -; |
| DR | EMBL; AF008749; AAL20073.1; -; |
| DR | EMBL; U43280; AAC43598.1; -; |

DR PIR; JC6040; JC6040.
DR StyGene; SGI0609; csgB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL : 1 21 POTENTIAL.

| FT CHAIN | 22 | 151 | MINOR CURLIN SUBUNIT. |
|-------------|---|--------------------------------------|-----------------------|
| SQ SEQUENCE | 151 AA; 16182 MW; C0FC5430E6DD361D CRC64; | | |
| Query Match | | 13 0%. Score 99.5. DB 1. Length 151. | |

Best Local Similarity 28.2%; Pred. No. 0.088;
Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

05 NCBENINATYDOL VTBIBIATYTHMAUEN CONNA LUNOATACDCCIMATDVCIGCENINAMAYOV 157

QY 99 NOFKNNAITLQVIRKVIHETMAGGNNAAVNIQASUSSVMVKQVGFNNNAIWWI 131
74 EGGNPRKVDQAGNINFAI-IEQTGMANDASISQSAHGNSAIIQSGGNKANITQY 129
Db

RESULT 7
OVO DROME
ID - OVO DROME
STANDARD;
PRT; 1028 AA.

| | |
|----|---|
| AC | FBI3211-09AZ04; |
| DT | 01-OCT-1996 (Rel. 34, Created) |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update) |

DE OVO protein (Shaven baby protein)
GN OVO OR SVB.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

| | |
|----|--|
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; |
| OC | Ephydroidea; Drosophilidae; Drosophila. |
| OX | NCBI_TaxID=7227; |
| RN | [1]_ |

| | |
|----|---|
| RP | SEQUENCE FROM N.A. |
| RC | TISSUE=Ovary; |
| RX | MEDLINE=95021209; PubMed=7935398; |
| RA | Garfinkel N.D., Wang J., Liang Y., Mahowald A.P.; |

RT "Multiple products from the shavenbaby-ovo gene region of *Drosophila*
RT melanogaster: relationship to genetic complexity.";
RL Mol. Cell. Biol. 14:6809-6818(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-Oregon-R;
RX MEDLINE=91293102; PubMed=1712294;
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;

RT The ovo gene of *Drosophila* encodes a zinc finger protein required
RT for female germ line development." ;
REL EMBO J., 10:2259-2266(1991).
CC - !- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM

| | |
|----|---|
| CC | LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION. |
| CC | -!- SUBCELLULAR LOCATION: Nuclear (Potential). |
| CC | -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND ACCUMULATES IN NURSE CELLS DURING OOCYGENESIS STOPPED IN THE EGG |

CC BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U11383; AAB60216.1; -
EMBL; X59772; CAB36921.1; ALT_SEQ.

PIR; A56038; A56038.

HSSP; P07248; 2ADR.

TRANSFAC; T00669; -

Flybase; FBgn0003028; ovo.

InterPro; IPR007087; Znf_C2H2.

Pfam; PF00096; zf-C2H2; 3.

SMART; SM00355; Znf_C2H2; 4.

PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.

PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.

Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;

Transcription regulation.

DOMAIN 62 66 POLY-ALA.

DOMAIN 72 77 POLY-GLY.

DOMAIN 80 85 POLY-GLY.

DOMAIN 98 108 POLY-GLY.

DOMAIN 144 152 POLY-HIS.

DOMAIN 153 159 POLY-ASN.

DOMAIN 336 339 POLY-GLN.

DOMAIN 347 353 POLY-GLN.

DOMAIN 357 361 POLY-GLN.

DOMAIN 410 414 POLY-GLN.

DOMAIN 422 426 POLY-GLN.

DOMAIN 426 432 POLY-GLN.

DOMAIN 445 453 POLY-GLN.

DOMAIN 456 459 POLY-GLN.

DOMAIN 466 474 POLY-GLN.

DOMAIN 497 517 POLY-ALA.

DOMAIN 524 529 POLY-SER.

DOMAIN 549 558 POLY-ALA.

DOMAIN 639 651 POLY-ALA.

DOMAIN 717 725 POLY-ALA.

DOMAIN 797 802 POLY-GLN.

DOMAIN 820 823 POLY-GLN.

DOMAIN 826 832 POLY-GLN.

DOMAIN 874 896 C2H2-TYPE 1.

DOMAIN 902 924 C2H2-TYPE 2.

DOMAIN 930 953 C2H2-TYPE 3.

DOMAIN 969 992 C2H2-TYPE 4.

DOMAIN 647 677 A -> R (IN REF. 2).

CONFLICT 647 677 A -> R (IN REF. 2);

SEQUENCE 1028 AA; D7068BB2BC0F677 CRC64;

Query Match 12.0%; Score 92.5; DB 1; Length 1028;

Best Local Similarity 26.0%; Pred. No. 2.9;

Matches 38; Conservative 15; Mismatches 66; Indels 27; Gaps 4;

3 LKVAFAAIVVSGALAGVFWQGGNGHNGGNSGPDSTLSIYQVGSANAALQSD 62

59 LQNAAAAYVINGAGS-----GGCTGGGAGSGPGGGSANSGGGGGG----- 104

63 ARKSETTITQSGYGVGADYGVGGADNSITELTNGFRNNATYDQLVTRVVTHEMAFAGNN 122

105 -----GGNGYVCGGVG-CPNNSLDGNNLLNFAVSYNESKSFNNHHHHQHNNN 155

123 AALVNTASDSSVMVRQVGFNNATA 148

156 ----NNNNGGQTSMMGHPYGGNPSA 177

RESULT 8

TNKL_HUMAN

ID TNKL_HUMAN STANDARD; PRT; 1327 AA.

AC O95271; O95272;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase 1) (TNKS-1) (TRF1-
DE interacting ankyrin-related ADP-ribose polymerase).
GN TNKS OR TNKS1 OR TINI OR TINI1 OR PARPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Testis;
RX MEDLINE=99040105; PubMed=9822378;
RA Smith S., Giriat I., Schmitt A., de Lange T.;
RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
RL Science 282:1484-1487(1998).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99454782; PubMed=10523501;
RA Smith S., de Lange T.;
RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
RT to nuclear pore complexes and centrosomes.";
RL J. Cell Sci. 112:3649-3656(1999).
RN [3]
RP FUNCTION, AND PHOSPHORYLATION.
RX MEDLINE=20556282; PubMed=10988299;
RA Chi N.-W., Lodish H.F.;
RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
RT substrate that interacts with IRAP in GLUT4 vesicles.";
RL J. Biol. Chem. 275:38437-38444(2000).
RN [4]
RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
RX MEDLINE=21602874; PubMed=11739745;
RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
RT at human telomeres.";
RL Mol. Cell. Biol. 22:332-342(2002).
CC -!- FUNCTION: May regulate vesicle trafficking and modulate the
CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
CC activity and can modify TRF1, and thereby contribute to the
CC regulation of telomere length.
CC -!- CATALYTIC ACTIVITY: NAD(+) + (ADP-D-riboseyl)(N)-acceptor =
CC nicotinamide + (ADP-D-riboseyl)(N+1)-acceptor.
CC -!- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
CC the cytoplasmic domain of INPP5/Oase in SLC2A4/GLUT4-vesicles.
CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is
CC also found at nuclear pore complexes and around the pericentriolar
CC matrix of mitotic centrosomes. During interphase, a small fraction
CC of TNKS is found in the nucleus, associated with TRF1.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O95271-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95271-2; Sequence=VSP_004538; VSP_004539;
CC Note=No experimental confirmation available;
CC TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues
CC by MAPK kinases.
CC -!- PTM: ADP-ribosylated (-auto).
CC -!- SIMILARITY: Belongs to the PARP family.
CC -!- SIMILARITY: Contains 15 ANK repeats.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; AF082556; AAC79841.1; -

```

DR EMBL; AF082557; AAC79842.1; -.
DR EMBL; AF082558; AAC79843.1; -.
DR EMBL; AF082559; AAC79844.1; -.
DR HSSP; Q00420; IAWC.
DR MIM; 603303; -.
DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
DR GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00023; ank; 19.
DR Pfam; PF00536; SAM; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 17.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS50088; ANK_REPEAT; 15.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
KW Phosphorylation; Alternative splicing.
FT REPEAT 215 247
FT REPEAT 248 280
FT REPEAT 281 313
FT REPEAT 368 400
FT REPEAT 401 433
FT REPEAT 434 466
FT REPEAT 521 556
FT REPEAT 557 589
FT REPEAT 590 622
FT REPEAT 683 715
FT REPEAT 716 748
FT REPEAT 749 781
FT REPEAT 836 868
FT REPEAT 869 901
FT REPEAT 902 934
FT DOMAIN 1030 1089
FT DOMAIN 1176 1327
FT DOMAIN 9 14
FT DOMAIN 27 34
FT DOMAIN 128 134
FT DOMAIN 137 145
FT VARSPLIC 641 643
FT VARSPLIC 644 1327
FT MUTAGEN 1184 1184
FT MUTAGEN 1291 1291
FT SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;

Query Match 11.9%; Score 91.5; DB 1; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.7;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAPAAI-VVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQVGSANAALALQSDAR 64
Db 99 VAAPVAVNSTSAGAVANPAGSGNNPSSSSPTSS-SSSSSPSSSLAESPEAA 157

QY 65 KSETTIT----QSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120
Db 158 GVSSTAPLPGAGGFGTGPVAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206

QY 121 NNAALVNQTSADSSVMVRYQVFG 143
Db 207 VNAK--DMAGKSSPLHFAAGFG 227

RESULT 9
OMPB_RICJA

ID OMPB_RICJA STANDARD; PRT; 1656 AA.
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (188 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RA Uchiyama T.;
RT Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica."
RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
CC Pfam; PF03797; Autotransporter; 1.
CC TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 11.7%; Score 90; DB 1; Length 1656;
Best Local Similarity 23.5%; Pred. No. 8;
Matches 40; Conservative 21; Mismatches 57; Indels 52; Gaps 6;

QY 6 VAAPAAI-VVSGSALAGVVPQWGGGNHNGSGSPDSTLSIYQVGSANAALALQSDARK 65
Db 509 VLAAGAITLDGSAIT-----TGDIGNGG-----GAALQSITLANDATK 547

QY 66 SETTITQSG-----YNGADVGQADNSTIETL-----QNGFRNNATYD 104
Db 548 ----TTLGGANIIISANGGTINFQANGGTIKTISTNNIIVDCDLAIAITDQTGVDPASSLT 604

QY 105 QLVTRVVTHEMAHAGGNNAL-----VNQTSADSSVMVRYQVFGNNAT 147
Db 605 NQAQTLTISGTIGIICANNITLGGQFNIGSKTTLNGNVAINELVIGNNS 654

RESULT 10
NI00_YEAST
ID NI00_YEAST STANDARD; PRT; 959 AA.
AC Q02629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)

```


FT DISULFID 454 498 BY SIMILARITY.
 FT DISULFID 503 553 BY SIMILARITY.
 FT DISULFID 523 546 BY SIMILARITY.
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;

Query Match 11.3%; Score 86.5; DB 1; Length 590;
 Best Local Similarity 47.1%; Pred. No. 4.9;
 Matches 24; Conservative 4; Mismatches 18; Indels 5; Gaps 2;

QY 103 YDQLVTRVVTHEMAHAGNNAALVNQTASDSVM-VROVGFQ----NNATA 148
 |||||
 Db 241 YDQLVTRVVTHEMAHAGLGSVVFPRDARILEISNVRHKDFDPVINSSTA 291
 |||||

RESULT 14
 GP63 LEIME STANDARD; PRT; 646 AA.
 AC P43150;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
 GN GP63-C1.
 OS Leishmania mexicana.
 OC Eukaryota; Eulenczoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5665;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=MNYC/BZ/62/M379;
 RX MEDLINE=93149206; PubMed=8426614;
 RA Medina-Acosta E., Kares R.E., Russell D.G.;
 RT "Structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated.";
 RL Mol. Biochem. Parasitol. 57:31-46(1993).
 CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1', and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-|-Leu-Lys-Lys-
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the amastigote forms.
 CC -!- SIMILARITY: Belongs to peptidase family M8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X64394; CAA45733.1; -.
 CC F01916; S19916.
 CC HSP; P08148; 1LML.
 CC MEROPS; M08.001; -.
 CC GlycoSuiteDB; P43150; -.
 CC InterPro; IPR006025; Pept_M_Zn_BS.
 CC InterPro; IPR001577; Peptidase_M8.
 CC Pfam; PF01457; Peptidase_M8_1.
 CC PRINTS; PR00782; LSHMANOLYSIN.
 CC PROSITE; PS00442; ZINC_PROTEASE; 1.
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
 KW Zymogen; Signal; Cell adhesion; Multigene family.
 FT SIGNAL 1 39 POTENTIAL.
 FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 103 646 LEISHMANOLYSIN C1.
 FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 267 267 BY SIMILARITY.
 FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DISULFID 127 144 BY SIMILARITY.
 FT DISULFID 193 232 BY SIMILARITY.
 FT DISULFID 316 388 BY SIMILARITY.
 FT DISULFID 395 458 BY SIMILARITY.
 FT DISULFID 408 427 BY SIMILARITY.
 FT DISULFID 417 492 BY SIMILARITY.
 FT DISULFID 469 513 BY SIMILARITY.
 FT DISULFID 513 568 BY SIMILARITY.
 FT DISULFID 538 581 BY SIMILARITY.
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 646 AA; 69054 MW; FE48DDC78C10B0A CRC64;

Query Match 11.2%; Score 86; DB 1; Length 646;
 Best Local Similarity 94.4%; Pred. No. 6;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 103 YDQLVTRVVTHEMAHAGG 120
 |||||
 Db 256 YDQLVTRVVTHEMAHAGV 273
 |||||

RESULT 15
 ICEN_XANCT STANDARD; PRT; 1567 AA.
 AC P18127;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAX.
 OS Xanthomonas campestris (pv. translucens).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OX NCBI_TaxID=343;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=X56S;
 RX MEDLINE=91080859; PubMed=2259339;
 RA Zhao J., Orser C.S.;
 RT "Conserved repetition in the ice nucleation gene inax from Xanthomonas campestris pv. translucens.";
 RL Mol. Gen. Genet. 223:163-166(1990).
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate crystallization in supercooled water. (By similarity).
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52970; CAA37140.1; -.
 CC HSP; P06620; 1INA.

Search completed: August 2, 2004, 14:49:33
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds
(without alignments)
1604.150 Million cell updates/sec

Title: US-09-543-407-30
Perfect score: 768
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 672 | 87.5 | 152 | 2 | O33802 | O33802 salmonella |
| 2 | 570.5 | 74.3 | 150 | 2 | O7X243 | O7X243 citrobacter |
| 3 | 534 | 69.5 | 149 | 2 | O7X240 | O7X240 citrobacter |
| 4 | 506.5 | 66.0 | 152 | 16 | O8CW63 | O8CW63 escherichia |
| 5 | 435.5 | 56.7 | 150 | 2 | Q7X237 | Q7X237 enterobacte |
| 6 | 380 | 49.5 | 176 | 2 | Q34069 | Q34069 salmonella |
| 7 | 126 | 16.4 | 502 | 16 | O8EIH4 | O8EIH4 shewanella |
| 8 | 122 | 15.9 | 29 | 2 | Q8S3J5 | Q8S3J5 escherichia |
| 9 | 112 | 14.6 | 139 | 16 | O8EIH3 | O8EIH3 shewanella |
| 10 | 111.5 | 14.5 | 171 | 16 | O89JI3 | O89JI3 bradyrhizob |
| 11 | 105 | 13.7 | 130 | 16 | O89JI4 | O89JI4 bradyrhizob |
| 12 | 104.5 | 13.6 | 151 | 16 | Q7UCZ1 | Q7UCZ1 shigella fl |
| 13 | 104.5 | 13.6 | 160 | 16 | O8CW64 | O8CW64 escherichia |
| 14 | 104.5 | 13.6 | 160 | 16 | O83RU7 | O83RU7 shigella fl |
| 15 | 104 | 13.5 | 157 | 16 | O88HG0 | O88HG0 pseudomonas |
| 16 | 102.5 | 13.3 | 3659 | 16 | Q98LN6 | Q98LN6 rhizobium 1 |

| | | | | | | |
|----|-------|------|------|----|--------|---------------------|
| 17 | 102 | 13.3 | 1422 | 16 | Q8EFU3 | Q8efu3 shewanella |
| 18 | 101.5 | 13.2 | 531 | 13 | Q8AWA8 | Q8awa8 lampetra fl |
| 19 | 101 | 13.2 | 1748 | 5 | Q94821 | Q94821 tetrahymena |
| 20 | 101 | 13.2 | 7716 | 16 | Q7UWZ8 | Q7uwz8 rhodospirell |
| 21 | 99.5 | 13.0 | 1209 | 16 | O89CK5 | O89ck5 bradyrhizob |
| 22 | 99.5 | 13.0 | 2174 | 16 | Q92U08 | Q92u08 rhizobium m |
| 23 | 99 | 12.9 | 562 | 16 | Q89D03 | Q89d03 bradyrhizob |
| 24 | 99 | 12.9 | 441 | 2 | O87327 | O87327 pseudomonas |
| 25 | 98.5 | 12.8 | 348 | 13 | O93397 | O93397 cyprinus ca |
| 26 | 98.5 | 12.8 | 624 | 3 | Q8NIV1 | Q8nivi neurospora |
| 27 | 98 | 12.8 | 145 | 16 | Q8U6N9 | Q8u6n9 agrobacteri |
| 28 | 97.5 | 12.7 | 91 | 2 | Q9S3J8 | Q9s3j8 escherichia |
| 29 | 97.5 | 12.7 | 477 | 2 | Q7X4S5 | Q7x4s5 pseudomonas |
| 30 | 97.5 | 12.7 | 586 | 16 | Q8EXJ2 | O8exj2 leptospira |
| 31 | 97.5 | 12.7 | 845 | 16 | Q7UIC5 | Q7uic5 mycobacteri |
| 32 | 97.5 | 12.7 | 646 | 16 | O53818 | O53818 mycobacteri |
| 33 | 97 | 12.6 | 1410 | 16 | O8CMJ0 | Q8cmj0 shewanella |
| 34 | 96.5 | 12.6 | 191 | 10 | Q7XDR3 | Q7xdr3 oryza sativ |
| 35 | 96.5 | 12.6 | 313 | 10 | O9SYZ5 | O9syz5 arabidopsis |
| 36 | 95.5 | 12.4 | 151 | 2 | Q7X244 | Q7x244 citrobacter |
| 37 | 95.5 | 12.4 | 154 | 16 | Q89JI5 | Q89ji5 bradyrhizob |
| 38 | 95.5 | 12.4 | 346 | 5 | Q9SR55 | Q9srs5 drosophila |
| 39 | 95.5 | 12.4 | 908 | 5 | Q9VM71 | Q9vm71 drosophila |
| 40 | 95 | 12.4 | 151 | 2 | Q7X238 | Q7x238 enterobacte |
| 41 | 95 | 12.4 | 196 | 10 | O22638 | O22638 zea mays (m |
| 42 | 94.5 | 12.3 | 153 | 16 | O89JI6 | O89ji6 bradyrhizob |
| 43 | 94.5 | 12.3 | 908 | 5 | O9VBC4 | O9vbc4 drosophila |
| 44 | 94.5 | 12.3 | 909 | 5 | Q8T4E0 | Q8t4e0 drosophila |
| 45 | 94.5 | 12.3 | 1713 | 3 | Q8TGE1 | Q8tge1 saccharomyc |

ALIGNMENTS

RESULT 1

O33802 ID O33802 PRELIMINARY; PRT; 152 AA.

AC O33802;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE AgfA protein (Fragment).

GN AGFA.

OS Salmonella typhimurium.

CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI_TaxID:602;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE:98053981; PubMed:9393832;

RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RT cells."

RL Infect. Immun. 65:5320-5325 (1997).

DR EMBL; AJ000514; CA04151.1; -.

FT NON_TER 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.5%; Score 672; DB 2; Length 152;

Best Local Similarity 89.4%; Pred. No. 38-46; Mismatches 13; Indels 0; Gaps 0;

Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSAGVVPQWGGGNGNGGNSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITOSGYNGGADVCGGADNSTIETQGFRRNATYDQLVTVVTHEVAHAGG 120

Db 61 SPARKSETTITOSGYNGGADVCGGADNSTIETQGFRRNATYDQLVTVVTHEVAHAGG 120

QY 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151


```
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 56.7%; Score 435.5; DB 2; Length 150;
Best Local Similarity 60.3%; Pred. No. 2.2e-27;
Matches 91; Conservative 26; Mismatches 33; Indels 1; Gaps 1;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQMGNGNHNHGGNSGPDSTLTSTIYQYGSNAALALQ 60
Db 1 MKPIKVAALAAIVVSGSAGAGMINQ-CGWGHGHGGYGGPNSLTNIYQNGGNSALALQ 59

Qy 61 SDARKSETTTTQSYGNGADVGQADNSTLTQNGFRNNATYDQLVTRVVTHEMAHAG 120
Db 60 TDARNVNLNISTQGGNGADVGQDDSSINLTQNGFGNSATLDOWNKSDVMNVQSYGG 119

Qy 121 NNAALVNOTASDSVMVQVQFGNNATANQY 151
Db 120 LNALVDQTASNSTVNTVQTQFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069,
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON_TER 1 76
FT NON_TER 76
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.5%; Score 380; DB 2; Length 76;
Best Local Similarity 97.4%; Pred. No. 2.7e-23;
Matches 74; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 30 GNHNGGNSGPDSTLTSTIYQYGSNAALALQSDARKSETTTTQSYGNGADVGQADNST 89
Db 1 GNHNGGNSGPDSTLTSTIYQYGSNAALALQSDARKSETTTTQSYGNGADVGQADNST 60

Qy 90 IELTQNGFRNNATYDQ 105
Db 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4,
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE S00865.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
```

```
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heideberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Ramathavan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
RA Fiedblum T.V., Smith H.O., Venter J.C., Nealsen K.H., Fraser B.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AF015532; AAN53941.1; -.
DR TIGR; S00865; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 16.4%; Score 126; DB 16; Length 502;
Best Local Similarity 27.0%; Pred. No. 0.045;
Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

Qy 29 GGNHNG-----GGN-----SSGPDSTLTSTIYQYGSANA---ALALQS 61
Db 231 GDNHTGFYVALAGSENDISMEQEGSNNTAYLSMTTGDDNTVDITQDGSNTVGDLSIADI 290

Qy 62 DARKSETTTTQSYGNGADVGQADNSTLTQNGFRNNATYDQLVTRVVTHEMAHAGN 121
Db 291 QGDDNDITIKQKGDGSGAEFQWGDSDVDLQKQGDANFATFGAYGTD-NDFDLSKGDN 349

Qy 122 NNAALVNOTASDSVMVQVQFGN-----NATAN 149
Db 350 NELVAFATGEDNSIEISQEGDANFAYVDATGN 381

RESULT 8
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5,
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curlin of Escherichia coli O78:k80 isolates associated with
RT IS1 insert on in csGB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.9%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0035;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQMG 29
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGG 29

RESULT 9
```

Query Match 14.5%; Score 111.5; DB 16; Length 171;
Best Local Similarity 24.8%; Pred. No. 0.19; 54; Indels 43; Gaps 5;
Matches 40; Conservative 24; Mismatches 24; Indels 43; Gaps 5;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGFDSTLSIYQVGSANAALQ 60
DB 40 MKKJFFASVAVALSSAAQA-----NTSITVQVLVNGSSVTQ 78

QY 61 SDARKSETTITQSYGNGADVGQAD-----NSTIELTQNGFRNATYDQLVTRVTHEMA 116
DB 79 NGLTNDSSSTTQIGILNGASTMQGTSSFLNNYSTVQAGVQNSATTGV-----128

QY 117 HAGGNAALVNQTA-----SDSSVMVROVQVGF-NNATANQ 150
DB 129 -AFGNNGSAITQNSFGFPALQNNNSASVGQLSFGINTSVSQ 168

RESULT 11
Q89J14 PRELIMINARY; PRT; 130 AA.
ID Q89J14
AC Q89J14; 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE BL15299 protein.
GN BL15299
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 13.7%; Score 105; DB 16; Length 130;
Best Local Similarity 24.7%; Pred. No. 0.46;
Matches 37; Conservative 27; Mismatches 50; Indels 26; Gaps 4;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGFDSTLSIYQVGSANAALQSDA 63
DB 1 MRLTYLVATAALASALTATVDAQ-----AGNSA-----SVLQFGTNSFSIQTGS 45

QY 64 RKSETTITQSYGNGADVGQADNSTIELT-----QNGFRNATYDQLVTRVTHEMAHAG 120
DB 46 TSNNAATLQFGATNTATTLTGSLITVNTAVTCGGTATASALTATGV-----GG 97

QY 121 NNAALVNQTAASDSSVMVROVQVGFNNATANQ 150
DB 98 SNSSILGIGIGANTAGVQGLGILNGSTILQ 127

RESULT 12
Q7UCZ1 PRELIMINARY; PRT; 151 AA.
ID Q7UCZ1
AC Q7UCZ1;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

| | | |
|-----------|---|--|
| OC | Enterobacteriaceae; Shigella. | |
| OX | NCBI_TaxID=623; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=2457T / ATCC 700930 / Serotype 2a; | |
| RX | MEDLINE=22590274; PubMed=12704152; | |
| RA | Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W., | |
| RA | Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A., | |
| RA | Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S., | |
| RA | Schwartz D.C., Blattner F.R.; | |
| RT | "Complete genome sequence and comparative genomics of Shigella | |
| RT | flexneri serotype 2a strain 2457T."; | |
| RL | Infect. Immun. 71:2775-2786(2003). | |
| DR | EWBL; AE016981; AAP16542.1; - | |
| SQ | SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64; | |
| | Query Match 13.6%; Score 104.5; DB 16; Length 151; | |
| | Best Local Similarity 29.1%; Pred. No. 0.6; | |
| | Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3; | |
| QY | 38 SSGPDSLTSLSYQYGSANAALALQSDARKSETTITQSGYNGCADVGQAGADNSTIELTQNGF 97 | |
| Db | 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIQAGTNNNSAQLRQSGSKLLAVVAQEGS 76 | |
| QY | 98 RNNATYDQLVTRVVVTHEMAH---AGGNNAAALVNQTASDSSVWVRQVQFGNNATANQY 151 | |
| Db | 77 SNRAKIDQ----TGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129 | |
| RESULT 13 | | |
| Q8CW64 | | |
| ID | Q8CW64 PRELIMINARY; PRT; 160 AA. | |
| AC | Q8CW64 | |
| DT | 01-MAR-2003 (T-EMBLrel. 23, Created) | |
| DT | 01-MAR-2003 (T-EMBLrel. 23, Last sequence update) | |
| DT | 01-MAR-2003 (T-EMBLrel. 23, Last annotation update) | |
| DE | Minor curlin subunit precursor. | |
| DE | CSGB OR C1305. | |
| GN | Escherichia coli O6. | |
| OS | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | |
| OC | Enterobacteriaceae; Escherichia. | |
| OX | NCBI_TaxID=217992; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | STRAIN=O6:H1 / CFT073 / ATCC 700928; | |
| RX | MEDLINE=22388234; PubMed=12471157; | |
| RA | Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P., | |
| RA | Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D., | |
| RA | Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T., | |
| RA | Mobley H.L.T., Donnenberg M.S., Blattner F.R.; | |
| RT | "Extensive mosaic structure revealed by the complete genome sequence | |
| RT | of uropathogenic Escherichia coli."; | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002). | |
| DR | EWBL; AE016759; AAN79778.1; - | |
| KW | Complete proteome. | |
| SQ | SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64; | |
| | Query Match 13.6%; Score 104.5; DB 16; Length 160; | |
| | Best Local Similarity 29.1%; Pred. No. 0.65; | |
| | Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3; | |
| QY | 38 SSGPDSLTSLSYQYGSANAALALQSDARKSETTITQSGYNGCADVGQAGADNSTIELTQNGF 97 | |
| Db | 30 AAGYDLANSEYNF----AVNELSKSFNQAAIIQAGTNNNSAQLRQSGSKLLTVVAQEGS 85 | |
| QY | 98 RNNATYDQLVTRVVVTHEMAH---AGGNNAAALVNQTASDSSVWVRQVQFGNNATANQY 151 | |
| Db | 86 SNRAKIDQ----TGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138 | |
| RESULT 14 | | |
| Q83RU7 | | |
| Q83RU7 | PRELIMINARY; PRT; 160 AA. | |

Aug 3 10:54:47 2004

| | | | | |
|-----------------------|--------|--|----------------|--------------------|
| Query Match | 13.5%; | Score 104; | DB 16; | Length 157; |
| Best Local Similarity | 26.4%; | Pred. No. 0.69; | | |
| Matches | 37; | Conservative 20; | Mismatches 57; | Indels 26; Gaps 4; |
| QY | 12 | IIVSGSALAGVWPQWGGGHHNGNSGPDSTLSIIQYGSANALALQSDARKSETTIT | 71 | |
| DB | 43 | LLPAGAQAAVIECQGGJG--NRAALDQCGALLGRIVQSGAQRAYILQ-EGSGLVAIIS | 99 | |
| QY | 72 | QSGYGCNGADYVCGQADNSTIELTQNGPRNNATYDQLVTRVWTHEAHAGGNNALVNQTAS | 131 | |
| DB | 100 | QQGNGNSASI-----RQSGSSNNAITEQI-----GNDNSASIVQSGS | 136 | |
| QY | 132 | DSSVWFVQVGFGNNTANQY | 151 | |
| DB | 137 | GLNSSVTQAGNGQHVOITQY | 156 | |

Search completed: August 2, 2004, 14:54:42
Job time : 29.7 secs